United States Environmental Protection Agency
EPA Science Advisory Board (1400A) Washington, D.C.
EPA-SAB-EC-WKSHP-02-001 January 2002 www.epa.gov/sab

Workshop on the Benefits of Reductions in Exposure to Hazardous Air Pollutants: Developing Best Estimates of Dose-Response Functions

An SAB Workshop Report of an EPA/SAB Workshop

NOTICE

This report has been written as part of the activities of the EPA Science Advisory Board, a public advisory group providing extramural scientific information and advice to the Administrator and other officials of the Environmental Protection Agency. This workshop report, however, does not document an advisory committee meeting, organized with the purpose of providing advice to the Agency. Instead, it documents a workshop open to the public and announced in the FEDERAL REGISTER. This report has not been reviewed for approval by the Agency and, hence, the contents of this report do not necessarily represent the views and policies of the Environmental Protection Agency, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names or commercial products constitute a recommendation for use.

Distribution and Availability: This EPA Science Advisory Board report is provided to the EPA Administrator, senior Agency management, appropriate program staff, interested members of the public, and is posted on the SAB website (www.epa.gov/sab). Information on its availability is also provided in the SAB's monthly newsletter (*Happenings at the Science Advisory Board*). Additional copies and further information are available from the SAB Staff at the following address: US EPA Science Advisory Board (1400A), 1200 Pennsylvania Avenue, NW, US Environmental Protection Agency, Washington, DC 20460.

TABLE OF CONTENTS

1. BACKGROUND

Hazardous Air Pollutants (HAPs) include pollutants identified in the Clean Air Act; pollutants known to cause or suspected of causing cancer or other serious human health effects, such as birth defects, neurological damage, and respiratory disease.[1] While it is clear that reducing emissions of HAPs to the atmosphere will reduce exposure levels to chemical agents that can cause serious health problems, there is no scientific consensus on the best way to quantify such risk reductions for the purpose of a benefits analysis. Nor is there consensus on how to value the reductions in risk, or what risk measures will be needed for valuation. The goal of this workshop was to consider improvements to methods for estimating changes in health risks resulting from regulations of HAPs that can be combined with valuation functions to estimate monetized benefits of HAP reductions. Improved methods for assessing HAPs benefits will assist the Agency in analyzing the economic value of its programs and in preparing reports to Congress, such as analyses of the benefits and costs of the Clean Air Act (CAA), as required by Section 812 of the CAA.

The workshop responded to a recommendation from the Health and Ecological Effects Subcommittee of the Advisory Council on Clean Air Compliance Analysis in 1999. Members and consultants from the Science Advisory Board (SAB) served individually on a committee with Agency staff to plan the workshop, where Agency staff, SAB members and consultants, experts outside the Board, and the public explored possible new methods for monetizing HAPs benefits. EPA explicitly sought a broad spectrum of views at the workshop and did not seek a consensus recommendation from workshop participants.

The workshop brought together expert discussants in the fields of economics, health science, and risk assessment as related to managing HAPs, with the help of the workshop chair and moderator, Dr. Michael Kleinman, College of Medicine, University of California, Irvine, California. The workshop took a "case study" approach to address two main issues:

a) Whether it is possible to produce best estimates of the central tendencies and distributions of the dose-response functions for a set of well-defined health endpoints for each of the case-study HAPs for use in the future activities on air quality and exposure modeling and how that might best be done.

b) How best to identify limitations and uncertainties in both risk assessment methods and economic models with regard to changes in health risks from reductions in air toxic emissions.

---

[1] Section 112(b) of the Clean Air Act describes hazardous air pollutants as those pollutants "which present, or may present, through inhalation or other routes of exposure, a threat of adverse human health effects (including, but not limited to, substances which are known to be, or may reasonably be anticipated to be, carcinogenic, mutagenic, teratogenic, neurotoxic, which cause reproductive dysfunction, or which are acutely or chronically toxic) or adverse environmental effects"

Details about the workshop process, a list of participants and agenda can be found in Appendices A-B.

The centerpiece of the workshop was a dialogue between economists and risk assessors. Dr. Lester Lave, an economist from Carnegie-Mellon University, set the stage with a white paper that documented his view of what benefits assessors need to know; why current toxicological information is not suited to provide this information; and a research agenda that would improve this estimation.

In response to his challenge, three distinguished experts prepared white paper case studies (see Appendix F). Examining chemicals for which there are substantial databases on health effects and exposure, the case studies were designed to illustrate the diversity of situations among various HAPs:

a) Benzene: a case with substantial data on cancer and non-cancer effects and strong hazard identification.[2] Case presented by Dr. Bernard Goldstein, Environmental and Occupational Health Sciences Institute, Robert Wood Johnson School of Medicine.

b) Perchloroethylene: a case with substantial, but not as consistent, data on cancer and data on noncancer hazards.[3] Case presented by Dr. Lorenz Rhomberg, Gradient Corporation.

c) Manganese: a case of a neurotoxin that causes many structural and functional effects, including a condition similar to Parkinson's Disease, and whose effects may accelerate the process of aging or onset of disease, albeit perhaps significantly after exposure.[4] Case presented by Dr. Bernard Weiss, University of Rochester.

The White Papers can be found on the SAB website (www.epa.gov/sab) as Appendix F of this Workshop Report.

These presentations were followed by a discussion where expert panelists (three

[2] Although the first case (Benzene), exemplified a HAP for which there is substantial epidemiological and toxicological evidence, it also highlighted some common areas of uncertainty in HAP risk assessments: extrapolation from health effect observations at higher exposures to assessment of lower exposure scenarios (e.g., uncertainties regarding carcinogenesis and the shape of the cancer dose response curve at low exposures); issue of "background" levels; cumulative burden; and other exposure-related issues.

[3] The second white paper (Perchloroethylene) highlighted additional issues and raised the possibility that the conclusions drawn from toxicological databases may lead us to conflicting or inconclusive results. Specifically, uncertainties in extrapolating health effect observations from laboratory animals to humans were highlighted.

[4] The third paper (Manganese) explored a very different terrain. In this case, the author created a conceptual model highlighting the idea that specific studied endpoints may be markers for a broader set of conditions and that the challenge for economists may be valuing a shift in the onset of risk or in the risk profile/trajectory of a population over their lifetime (after perhaps a long latency period).

economists and three health scientists) each were asked to address the two central workshop questions and reflect on the approaches proposed in the three chemical-specific white papers. Most of the expert panelists, and the Workshop Co-Moderator, Dr. Roy Albert, expanded on their brief presentations in written comments submitted after the workshop. These written submissions can be found on the SAB website (www.epa.gov/sab) as Appendix G of this Workshop Report.

2. MAJOR THEMES

The three white papers and remarks by Dr. Lester Lave generated lively discussion but no consensus on a proposed methodology. Many specific options emerged in the discussion and in panelists' written comments. Appendix F lists many of the different strategies suggested for bridging the gaps. There was, however, general agreement that this was a fertile area for further study and that there was a need for further cross-disciplinary work between risk assessors and economists/analysts.

During the workshop discussion, it was noted that the different disciplines are pursuing fundamentally different questions. In general, regulatory toxicologists are asking, "What level is likely to be safe?" Economists, on the other hand, are asking, "What are the number of cases reduced per small unit change in pollution concentration from individual regulations?" Importantly, economists noted that they would like to have central tendencies and distributions, rather than upper-bound estimates.

Although the discussants were in general agreement that there were benefits associated with reductions in air toxics emissions, they were split in their opinion about whether using the criteria pollutant model would work for air toxics. The case studies stimulated discussion about the advisability of pursuing a pollutant-by-pollutant, endpoint-by-endpoint evaluation of economic benefits.

The difficulties raised by the first two case studies were discussed. Namely, for the 188 listed HAPs, if we followed the criteria pollutant model, some difficulties would include: limited health effect data; difficulties in modeling multiple assaults on a target organ; contradictory or inconclusive evidence with respect to how many cases might develop or what organ would be targeted in humans; difficulties in extrapolation from animal models; uncertainties in extrapolating to lower doses; difficulties in evaluating background exposures (especially for national policy applications); and lack of resources to evaluate fully each chemical or class of chemicals listed in the Clean Air Act. The inability of economists to value relatively small risk reductions from endpoints whose biological significance is difficult to understand (e.g., change in platelet count) would hamper this approach even if the biological data base were complete and yielded best estimates and a characterization of variability. It was also discussed that the purpose of the workshop was to discuss one area (dose-response estimates) but that there are uncertainties in all of the other analytical steps (e.g., emissions, dispersion, exposure, and valuation). See Dr. Cameron's written comments (Appendix G) for a fuller explanation.

Possible alternatives to the criteria pollutant approach were also suggested. One suggestion was to evaluate "cleaning up the dirty stuff" or thinking of the HAP regulation as an insurance policy. Thus, economists would study the utility or value of peace-of-mind derived from knowing that the public was protected from the bundle of effects, either known or unknown, related to the listed HAPs. Another suggestion was to evaluate whether the public

held a different value for full elimination of an involuntary risk versus small incremental changes in that risk. A third promising possibility was to study the value of avoiding entry into a risk pool or shifts in the curve of population's onset of disease (e.g., as suggested in Dr. Weiss's paper). The Agency is considering introducing approaches similar to these proposals as alternatives to a damage-function approach for assessing benefits for control of criteria pollutants.

Although there was no consensus among the discussants, the research directions suggested by this workshop suggest a dual approach. The first avenue would involve continuing to address the HAPs using a damage function approach of estimating health effects avoided by a given policy and determining the economic value of avoided effects. The first two case studies (benzene and percholoroethlyene) lay out these issues very clearly. In addition, the comments by Dr. Cameron and Dr. Zeise provide insights and research suggestions.

The second research approach might involve some of the other ideas discussed at the workshop, including the concept of valuing the bundle of HAP reduction efforts embodied in the Clean Air Act from an economic perspective (i.e., the utility or value of peace-of-mind derived from knowing that the public was protected from the bundle of effects, either known or unknown, as opposed to the avoided health effects), the approach discussed in the manganese case study, and the insurance concept suggested in Dr. Smith's comments.

One of the invited discussants, Dr. Bailar, summarized the situation aptly by saying that benefit-cost analysis is an information-hungry process which we apply to an information-sparse problem with respect to HAPs. Regulatory decision-making informed by benefit-cost analysis is a precision-hungry process that at present we base on precision-sparse inputs. The workshop provided a meaningful first step, and more discussion and collaboration between the risk assessment and benefits assessment communities is needed.

APPENDIX A: LIST OF WORKSHOP PARTICIPANTS

CHAIR AND MODERATOR

Dr. Michael T. Kleinman, College of Medicine, University of California, Irvine, CA.

CO-MODERATOR

Dr. Roy Alpert, Division of Environmental Health, University of Cincinnati, Cincinnati, OH.

KEY DISCUSSANTS

Dr. John C. Bailar III, Department of Health Studies, University of Chicago, Chicago, IL.

Dr. Trudy Cameron, Department of Economics, University of California, Los Angeles, CA.

Ms. Lauraine Chestnut, Stratus Consulting, Boulder CO.

Dr. James Cogliano, Office of Research and Development, U.S. Environmental Protection Agency, Washington, D.C.

Dr. James DeMocker, Office of Air and Radiation, U.S. Environmental Protection Agency, Washington, D.C.

Dr. A. Myrick Freeman, Department of Economics, Bowdoin College, Brunswick, ME.

Dr. Paul Locke, Pew Environmental Health Commission, Baltimore, MD.

Dr. Albert McGartland, Office of Policy, Economics and Innovations, U.S. Environmental Protection Agency, Washington, D.C.

Dr. Dennis Paustenbach, Exponent, Menlo Park, CA.

Dr. V. Kerry Smith, Center for Environmental and Resource Economics Policy, Department of Agricultural and Resource Economics, North Carolina State University
Raleigh, NC.

Dr. Jeanette Wiltse, Office of Water, U.S. Environmental Protection Agency, Washington, D.C.

Dr. Lauren Zeise , California Environmental Protection Agency, Oakland, CA.

PRESENTERS OF WHITE PAPERS

Dr. Bernard Goldstein, Environmental and Occupational Health Sciences Institute, Robert Wood Johnson School of Medicine, Rutgers, Piscataway, NJ.

Dr. Lester Lave, Graduate School of Industrial Administration, Carnegie-Mellon University Pittsburgh, PA.

Dr. Lorenz Rhomberg, Gradient Corporation, Cambridge, MA.

Dr. Bernard Weiss, University of Rochester, Rochester, NY.

SAB STAFF

Dr. Angela Nugent, Science Advisory Board, U.S. Environmental Protection Agency, Washington, D.C.

APPENDIX B: AGENDA

SAB/EPA Workshop on the Benefits of Reductions in Exposure to Hazardous Air Pollutants: Developing Best Estimates of Dose-Response Functions

June 22-23, 2000
Westin Grand Hotel
West 2350 M Street, NW, Washington DC

Welcome, Brief Overview of Workshop, and Introductions	Michael Kleinman, University Of California, Irvine
9:15. **Background**	
Current approaches to Cancer and Noncancer Risk Assessment	William Farland, ORD
History of cost/benefit analysis for air pollution in general for Section 812 of the Clean Air Act Amendments	Albert McGartland, EPA James DeMocker, EPA
10:30 am **Review of Workshop Scope and Purpose**	Michael Kleinman
10:45 am **Economist's Perspective on HAP Benefits Analysis Under Section 812**	Lester Lave, Carnegie Mellon
11:15 am **Discussion of Morning Papers**	Michael Kleinman, Moderator, Key Discussants All Attendees
12:00 Noon **Lunch**	
1:00 pm **Case Study Presentations** (For each: 20 minute risk assessor presentation of issues, 30 minute core panel discussion, 30 minute general discussion [comments from all attendees])	Roy Albert, University of Cincinnati, Discussion Moderator
1:00 pm **Benzene**	Bernard Goldstein, EOHSI, Rutgers Key Discussants All Attendees
2:20 pm **Perclorethylene**	Lorenz Rhomberg, Gradient Corporation Key Discussants All Attendees

4:00 pm
Manganese — Bernard Weiss, University of Rochester
Key Discussants
All Attendees

5:20 pm **Adjourn**

June 23, 2000

9:00
Recap of Day 1 and Review of Workshop Agenda for Day 2 — Michael Kleinman

9:15
Discussion of Questions Before the Workshop — Roy Albert and Michael Kleinman, Moderators
Key Discussants
All Attendees

(1) Proposed approaches for hazard assessments for selected HAPs that would facilitate benefit assessments for those chemicals — Lead Discussants: Paul Locke, Laurie Chestnut

(2) Expert discussants' views on whether it is possible to produce a methodology for developing central tendencies and distributions in hazard assessments for HAPs for use in benefits analyses and how that might best be done — Lead Discussants: Dennis Paustenbach, Rick Freeman

(3) How best to identify limitations and uncertainties in both risk assessment methods and economic models — Lead Discussants: John Bailar, Kerry Smith

(4) Suggestions and priorities for a research agenda to address identified gaps in available data and methods needed to conduct HAPs related benefit analyses — Lead Discussants: Lauren Zeise, Trudy Cameron

11:30 am
Summary and Identification of Next Steps — Michael Kleinman, Moderator

12:00
Adjourn

APPENDIX C: WORKSHOP PROSPECTUS DISTRIBUTED BEFORE THE WORKSHOP

SAB/EPA Workshop on the Benefits of Reductions in Exposure to Hazardous Air Pollutants: Developing Best Estimates of Dose-Response Functions

Purpose

Hazardous air pollutants (HAPs) have been the focus of a number of EPA regulatory actions, which have resulted in significant reductions in emissions of HAPs. EPA has been unable to adequately assess the economic benefits associated with health improvements from these HAP reductions due to a lack of best estimate dose-response functions for health endpoints associated with exposure to HAPs and also due to the air quality and exposure models for HAPs available for use in benefits analysis. EPA is conducting two activities to develop a proposed methodology to generate estimates of the quantified and monetized benefits of reductions in exposure to HAPs. The first will be a workshop focusing on developing best estimates of dose-response functions that relate changes in HAP exposure to changes in health outcomes. The second activity will focus on (1) integrating these dose-response functions with appropriate models of HAP concentrations and human exposure and (2) translating these into economic benefits that would estimate changes in health risks resulting from regulations that reduce HAP emissions.

The overall goal of these two activities is to identify methods for the Agency to consider using in estimating changes in health risks resulting from HAP regulations that can be combined with valuation functions to estimate monetized benefits of HAP reductions.

Risk assessments for HAPs have been developed to help decision makers set health-based standards that are consistent with EPA's mission to protect human health. The quantitative toxicity values from these assessments (that is, the cancer slope factors and the noncancer reference concentrations and reference doses) are typically based on animal and epidemiologic studies that involve higher exposures than those encountered in the environment. The gap between environmental doses and study doses has led to toxicity values that can put a bound on the actual risk without being able to provide a reliable central estimate or distribution of risks. It is these latter terms (central estimates and distributions) that economists have traditionally used to estimate the economic value of potential changes in risks.

In contrast, risk assessments for criteria pollutants have been based on epidemiologic and clinical studies of exposures similar to those encountered in the environment. This has allowed development of standard statistical confidence intervals and distributions. With this information, economists have been able to develop economic benefit estimates for many health endpoints related to criteria pollutants. Criteria pollutant benefit estimates have been feasible

because of the availability of: (a) well-defined health endpoints such as hospital admissions or premature mortality; (b) dose-response functions from epidemiological and clinical studies which support estimates of risk reductions in terms amenable to economic valuation; (c) reliable estimates of ambient concentration and population exposure change; and (d) dose-response functions available from epidemiological and clinical studies in which the exposures were similar to those being experienced in the ambient environment. Uncertainties related to the health benefits of criteria pollutants have generally been represented by standard confidence intervals based on measures of within and between study variation in the estimated health effects.

While mortality from HAP-related cancer is a well-defined endpoint, there are very few validated exposure-response relationships. For the many other potential health effects from exposure to HAPs, such as changes in reproductive functions or mutagenic effects, there are major information gaps in all aspects of risk assessment, as well as in exposure-response and valuation. The focus of this workshop will be the development of best-estimates and uncertainty characterizations for hazard and dose response functions for use in benefits analyses of HAP regulations, with a focus on providing potentially useful data and tools to support HAP-related benefit assessments, including national-scale program evaluations.

Expected outcomes from this workshop will include a report documenting: (1) proposed approaches for hazard assessments for selected HAPs that would facilitate benefit assessments for those chemicals; (2) expert discussants' views on whether it is possible to produce a methodology for developing central tendencies and distributions in hazard assessments for HAPs for use in benefits analyses and how that might best be done; (3) h ow best to identify limitations and uncertainties in both risk assessment methods and economic models; and (4) suggestions and priorities for a research agenda to address identified gaps in available data and methods needed to conduct HAPs related benefit analyses.

Scope

The workshop will task expert discussants, who have knowledge and expertise in the fields of economics, health science, and risk assessment as related to managing HAPs, to address the following issues:

I. Whether it is possible to produce best estimates of the central tendencies and distributions of the dose-response functions for a set of well-defined health endpoints for each of the case-study HAPs for use in the second activity on air quality and exposure modeling and how that might best be done.

II. How best to identify limitations and uncertainties in both risk assessment methods and economic models with regards to changes in health risks from reductions in air toxic emissions, especially in the following areas:

A. Defining and characterizing best estimates and uncertainty distributions for hazard and dose response functions for both cancer and non-cancer effects.

B. Defining the context for the use of conservative reference doses, including the basis and methodology of the risk assessment.

C. Identifying potential uncertainties from extrapolating effects from "high dose" occupational and animal exposure studies to lower level ambient HAP concentrations.

D. Incorporating currently/typically available toxicity (both human and animal) data sets into existing or modified economic benefit models.

E. Evaluating the usefulness of benefits models based on dose-response functions derived from epidemiological studies as models for HAP benefit analyses.

III. Identifying gaps in existing knowledge and developing a proposed research agenda.

APPENDIX D: STRATEGIES FOR BRIDGING THE GAPS BETWEEN ECONOMISTS AND HEALTH SCIENTISTS

The Workshop resulted in many suggestions to address the gaps between economists and health scientists to improve benefits assessments for HAPs. These suggestions included the following:

1. Analyze the implications of current *in vivo* and *in vitro tests* for human toxicity for all known human carcinogens. Compare laboratory results with human data; for chemicals that are positive in the NTP bioassay, identify which are and which are unlikely to be human carcinogens. (Lave)

2. Invest time and resources in developing expert judgment on the best estimate for a chemical, like benzene, with substantial data on cancer and non-cancer effects and strong hazard identification but limited data at low doses. (Goldstein)

3. Acknowledge that data sets have major uncertainties. Invest in statistical approaches to characterize the distribution of uncertainties, even for chemicals with significant uncertainty in their data sets. Approaches would establish explicit weights for different assumptions associated with estimation of risk. Approaches would allow examination of the contribution to overall uncertainty from various components. (Rhomberg)

4. Link endpoints in toxicology studies to adverse effects the public cares about. Communicate these linkages and associated uncertainties to economists conducting benefit assessments. (Weiss)

5. Estimate economic value of avoiding entry into a risk pool or shifts in the curve of population's onset of disease. (Weiss)

6. Multiply the estimated number of cancer cases by a weighting factor that is determined by the strength of the evidence. (Albert)

7. Educate congress, the public, and the news media about "the art of the possible" in risk assessment and the limits of science. (Bailar)

8. Establish closer links between risk analysts and cost-benefit analysts. Establish cross-training programs. Establish regular, weekly meetings on each project in which risk analysis will be a significant element of a cost-benefit analysis. (Bailar)

9. Analyze the level of accuracy needed at each step of analysis to make most effective use of limited resources. (Bailar)

10. Consider "bundling" or grouping HAPs that are similar, either according to health endoint, chemical species, biologic mechanism or mode of action, or source categories. (Bailar)

11. Keep in mind that there were uncertainties not only in dose-response estimates that feed into cost-benefit analysis, but also in all of other analytical steps (e.g., emissions, dispersion, exposure, and valuation). (Cameron)

12. Consider a "top down" approach to benefits assessment, derived from individuals' subjective assessments of what HAP regulation would be likely to achieve in terms of health effects (or other effects) instead of aiming for "an unambiguous, objectively calculated, bottom-up measure" calculated on a chemical-by-chemical basis. (Cameron)

13. Develop a benefits assessment based on a "better understanding of the relationship between controls on emissions of HAPs and individuals' perceptions of safety or 'peace of mind.'"(Freeman, Lave, Smith)

14. Sort HAPs into "bins" according to the adverse effects of concern, and conduct benefit-cost analyses for different "bins." Distinguish chemicals listed due to their carcinogenicity, from chemicals listed because they are systemic (non-carcinogenic) toxicants, from irritants. For each group, develop a dose-response curve and build a distribution of points around the dose-response curve (for both carcinogens and non-carcinogens). (Paustenbach)

15. Invest in research to improve understanding and quantitative descriptions of how risk may vary in a population. This research would produce better mean estimates of risk and clearer understanding of the magnitude of risk borne by different individuals. (Zeise)

Appendix E-1

"Current Approaches to Cancer and Noncancer Risk Assessment: Implications for Developing Best Estimates of Dose-Response Functions,"

Presented by Dr. William H. Farland

An SAB/EPA Workshop on the Benefits of Reductions
in Exposure to Hazardous Air Pollutants
June 22 and 23, 2000

Current Approaches to Cancer and Noncancer Risk Assessment:

Implications for Developing Best Estimates of Dose-Response Functions

William H. Farland, Ph.D., Director
National Center for Environmental Assessment
Office of Research and Development
U.S. ENVIRONMENTAL PROTECTION AGENCY

US EPA Office of Research and Development

National Center for Environmental Assessment

Recent Emphasis Focuses on the Use of *Mode-of-Action* Data

"The quality of risk analysis will improve as the quality of input improves. As we learn more about biology, chemistry, physics, and demography, we can make progressively better assessments of the risks involved. Risk assessment evolves continually, with reevaluation as new models and data become available."

"Science and Judgment in Risk Assessment" (National Research Council, 1994)

Breaking Down the Dichotomy

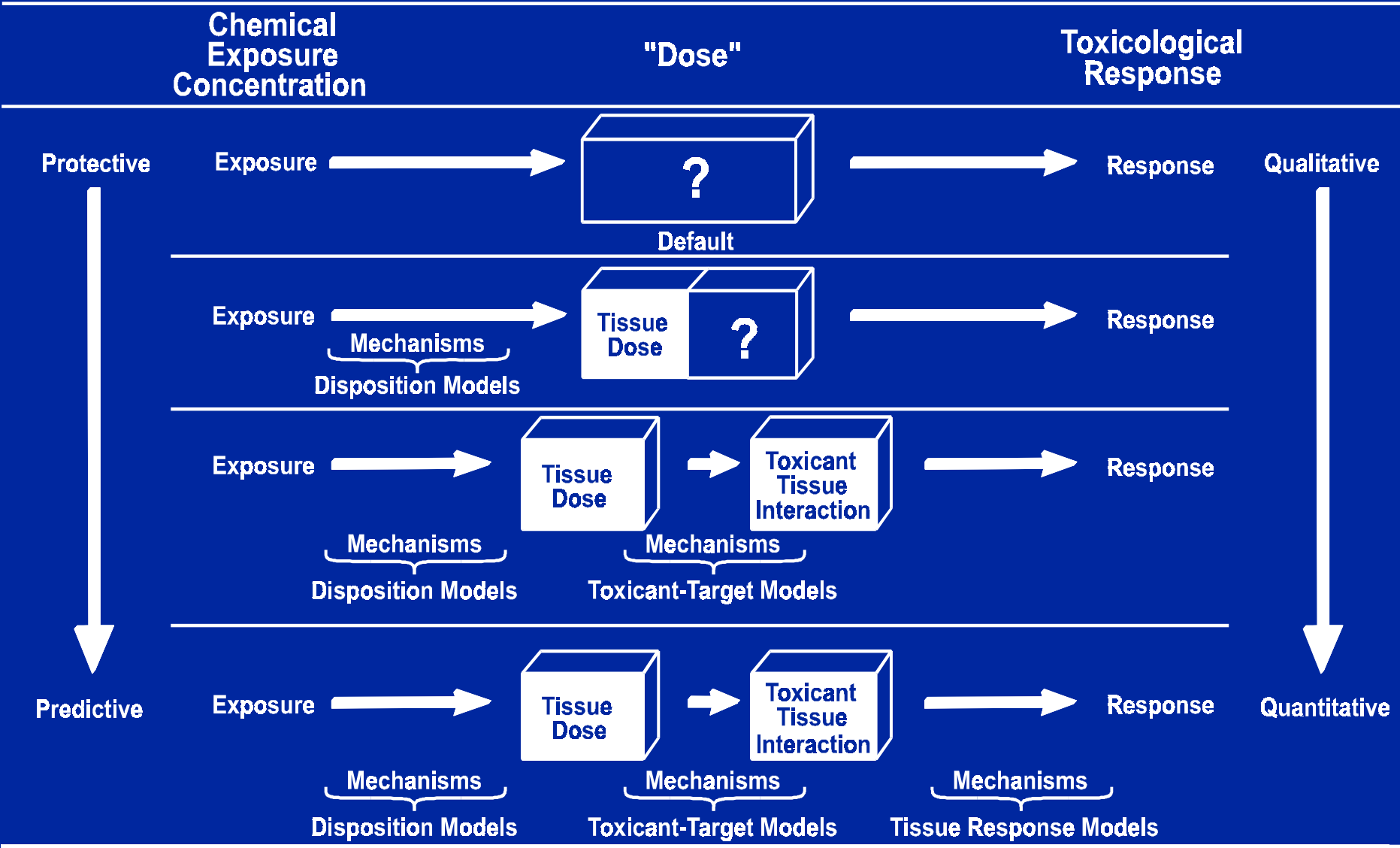
Cancer

- Non-Threshold
- Irreversible
- "Risk" value
  - Slope Factor
  - Unit Risk
  - Risk-Specific Dose

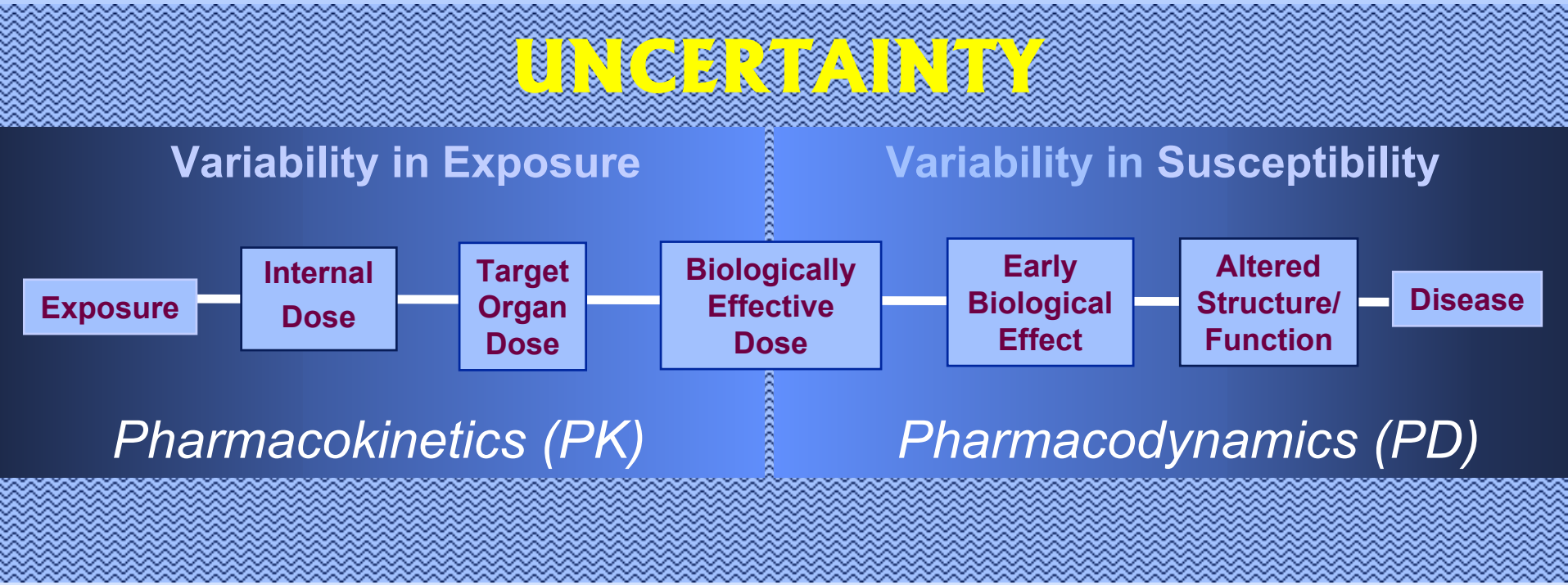
Non-Cancer

- Threshold
- Reversible
- "Safety" value
  - RfD/RfC
  - ADI/TDI
  - MRL

Systematic Characterization of Comprehensive Exposure-Dose-Response Continuum and the Evolution of Protective to Predictive Dose-Response Estimates



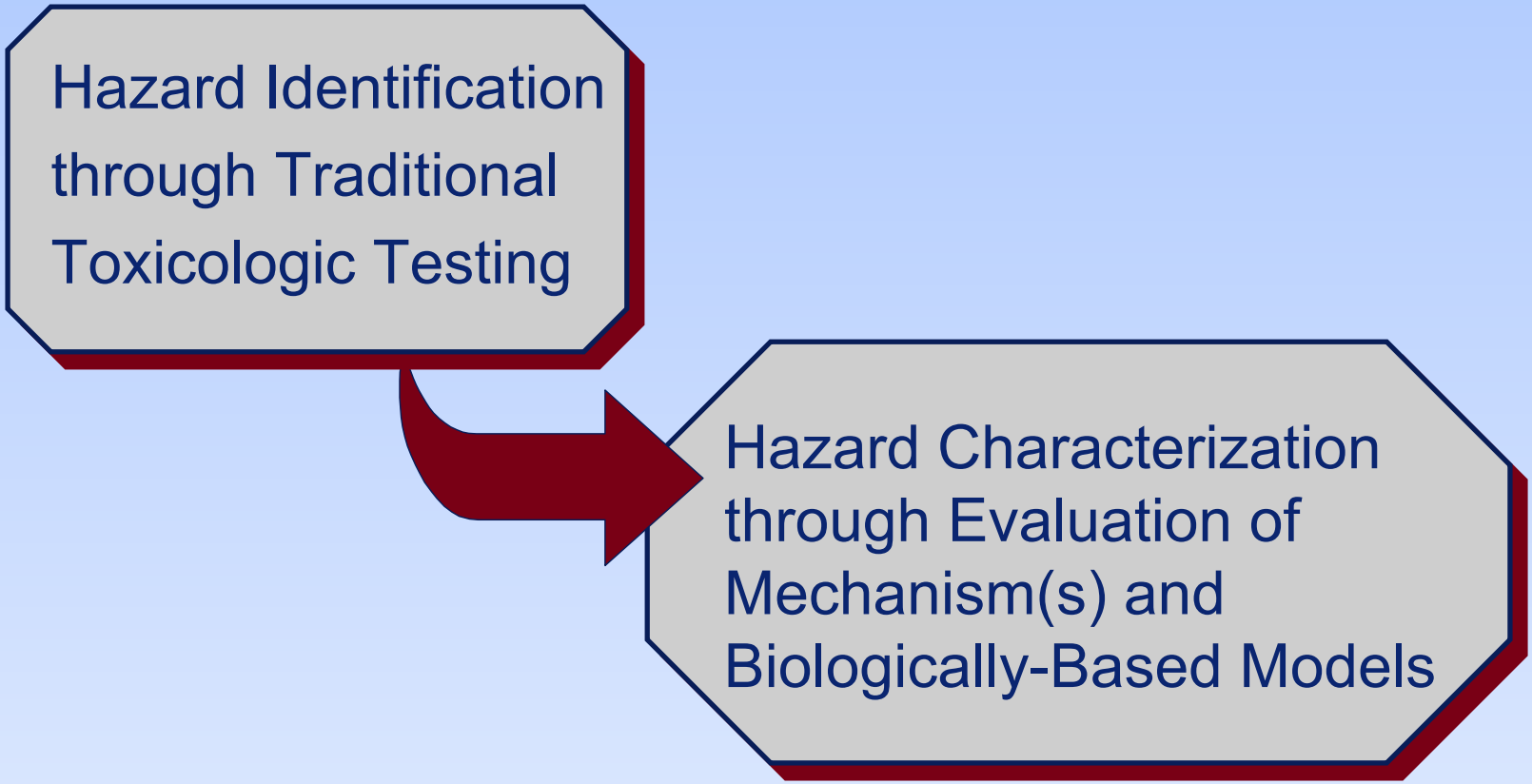
Uncertainty, Variability, and the Continuum Between Exposure and Disease



Revision Directions for Risk Assessment Guidelines --

- Emphasize full **characterization**
- Expand role of **mode-of-action** information (and, therefore, ***biomarkers !***)
- Use **all information** to design dose response approach
- **Two step** dose response assessment

Evolution of Hazard Characterization



BIOMARKERS --

Definition:

Biologic markers are indicators signaling events in biologic systems or samples.

Three types:

- ➔ Exposure
- ➔ Effect
- ➔ Susceptibility

Mechanism vs. Mode-of-Action

Mechanism of action:

Detailed molecular description of a key event in the induction of cancer or other health endpoints

Mode-of-Action:

Key events and processes, starting with the interaction of an agent with a cell, through functional and anatomical changes, resulting in cancer or other health endpoints

Mode-of-Action --

- How does the chemical produce its effect?
- Are there mechanistic data to support this hypothesis?
- Have other mechanistic hypotheses been considered and rejected?

How is mode-of-action information used?

Address Uncertainty in Risk Assessment:

- Comparative Structure Activity Relationships (SAR)
- Relevance of animal data for extrapolation
- Shape of dose-response curve
  - ➔ Range of Observation
  - ➔ Range of Inference
- Susceptibility of individuals/ subpopulations

Demonstrating a Mode-of-Action --

To show that a postulated ***mode-of-action*** is operative, it is generally necessary to:

- **outline** the sequence of events leading to effects;
- **identify** key events that can be measured; and
- **weigh** information to determine whether there is a causal relationship between events and cancer formation.

Framework --

- Summary Description of Postulated Mode-of-Action
- Topics:
  1. *"Identify key events" (→ BIOMARKERS**?**)*
  2. *"Strength, consistency, specificity of association"*
  3. *"Dose-response relationship"*
  4. *"Temporal relationship"*
  5. *"Biological plausibility and coherence"*
- Conclusion

Key Event --

Examples:

- Metabolism
- Receptor-ligand changes
- DNA or chromosome effects
- Gene transcription; protein synthesis
- Increased cell growth and organ weight
- Hormone or other physiological perturbations
- Hyperplasia, cellular proliferation

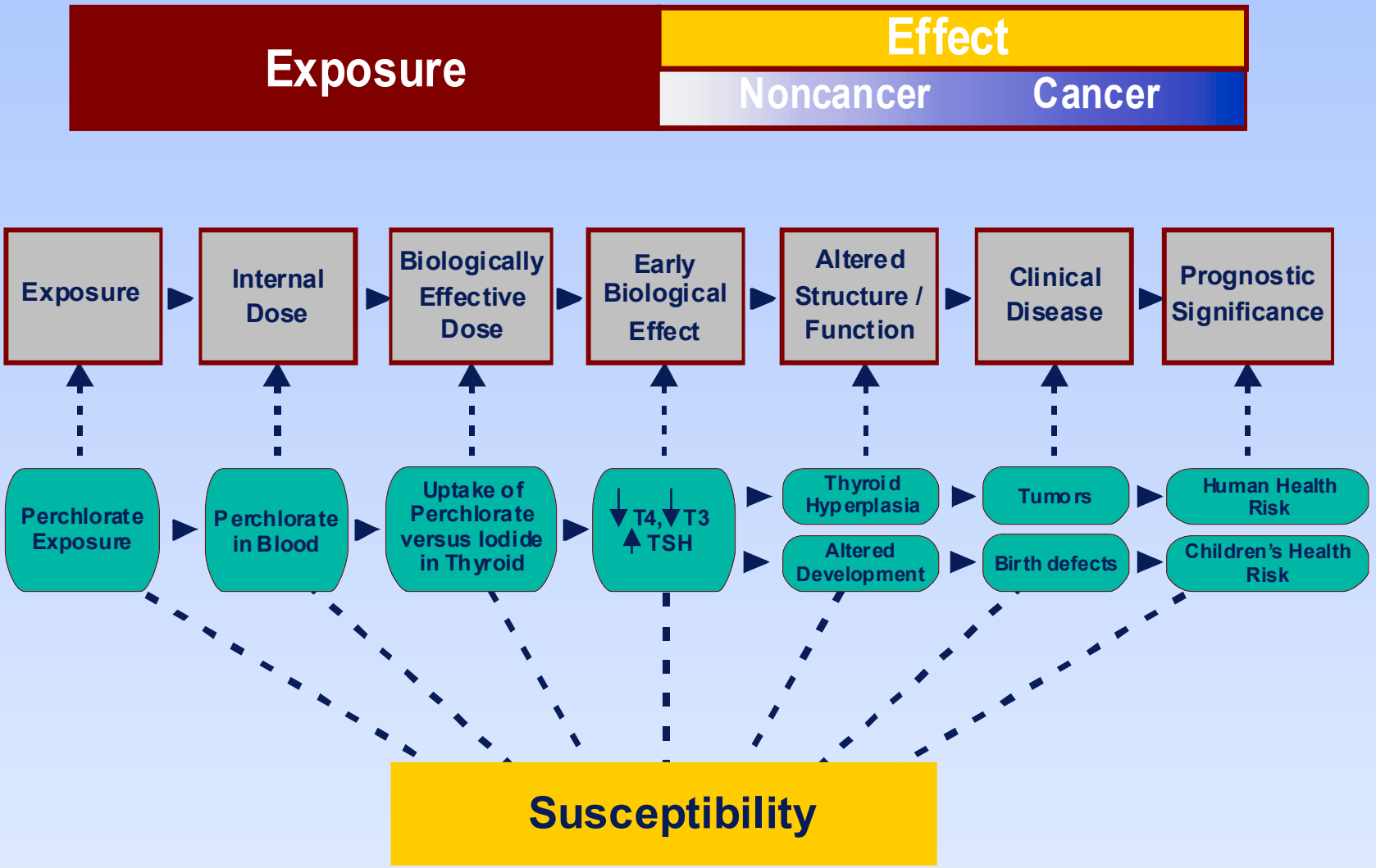
Use of Mode-of-Action Information: *Examples*

- Formaldehyde ⇒ DNA crosslinks ⇒ Cell proliferation
- Methylene Chloride ⇒ Pharmacokinetics ⇒ Genetic polymorphisms
- d-Limonene ⇒ "-2-u-globulin, etc.
- Chloroform ⇒ Cytotoxicity
- Dioxin ⇒ Receptor-mediated responses

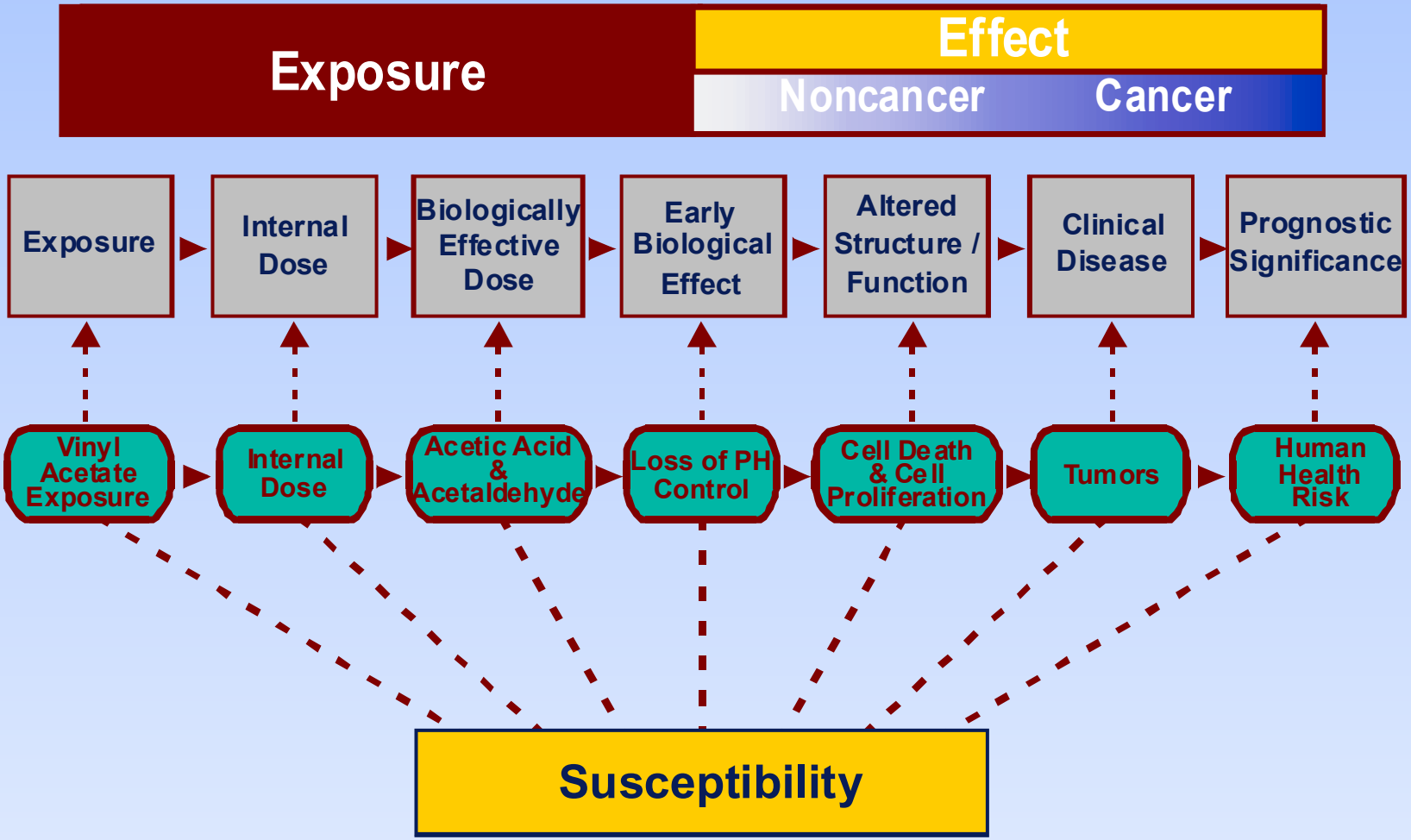
Use of Mode-of-Action Information: *More Examples*

- BaP ⇒ DNA reactive metabolites ⇒ Cell proliferation
- Amitrole ⇒ Increased Thyroid Stimulating Hormone (TSH) ⇒ Cell proliferation
- Melamine ⇒ Increased urinary pH ⇒ Irritation
- *Perchlorate* ⇒ *Altered thyroid homeostasis*
- *Vinyl Acetate* ⇒ *Cytotoxicity* ⇒ *Cell proliferation*

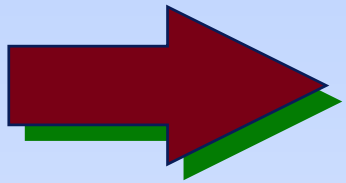
Proposed Mode-of-Action Model for Risk Assessment of Perchlorate



Proposed Mode-Of-Action Model for Risk Assessment of Vinyl Acetate



Mechanistic data refines interpretation and extrapolation of :

Exposure  Dose

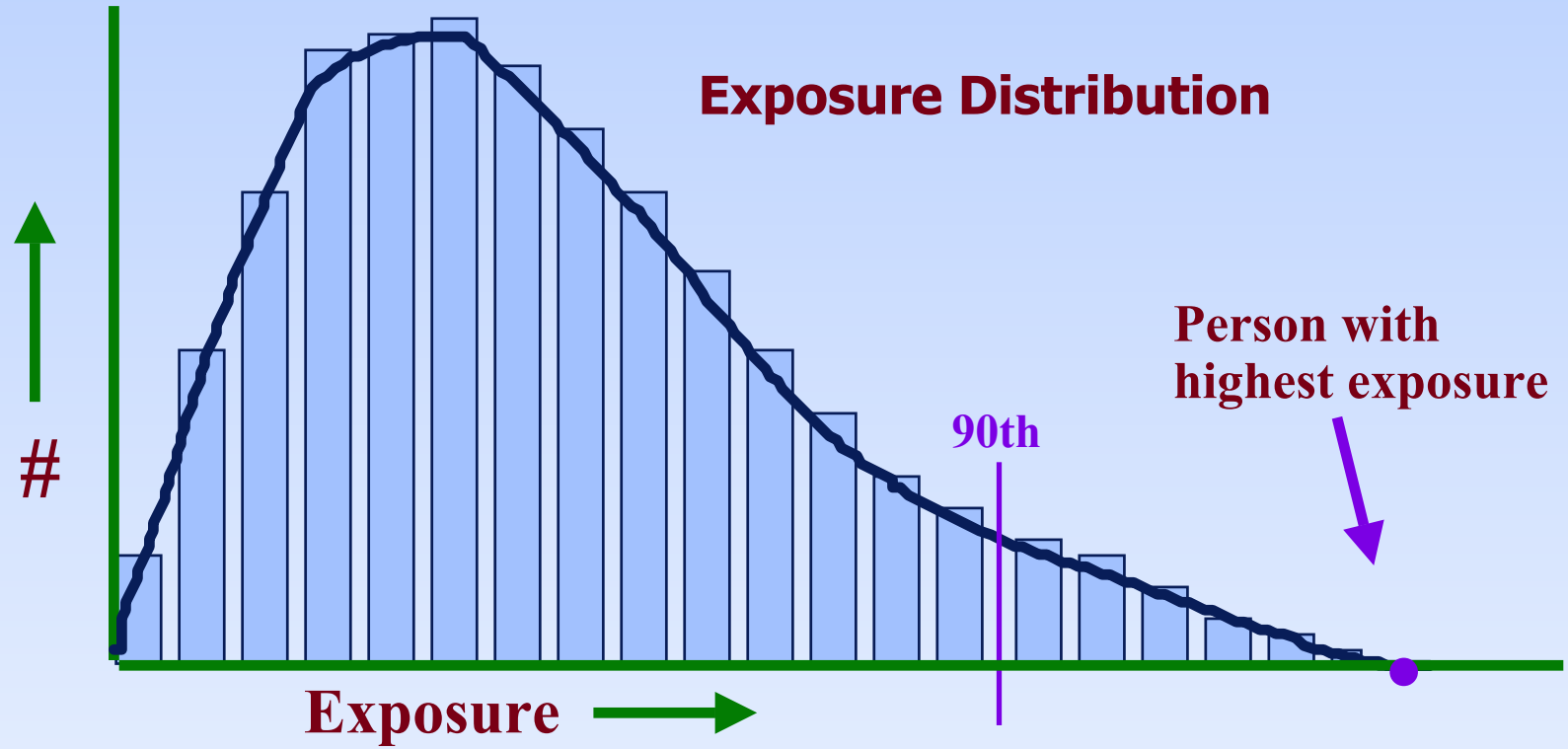
Relationships of Exposure and Dose to Risk

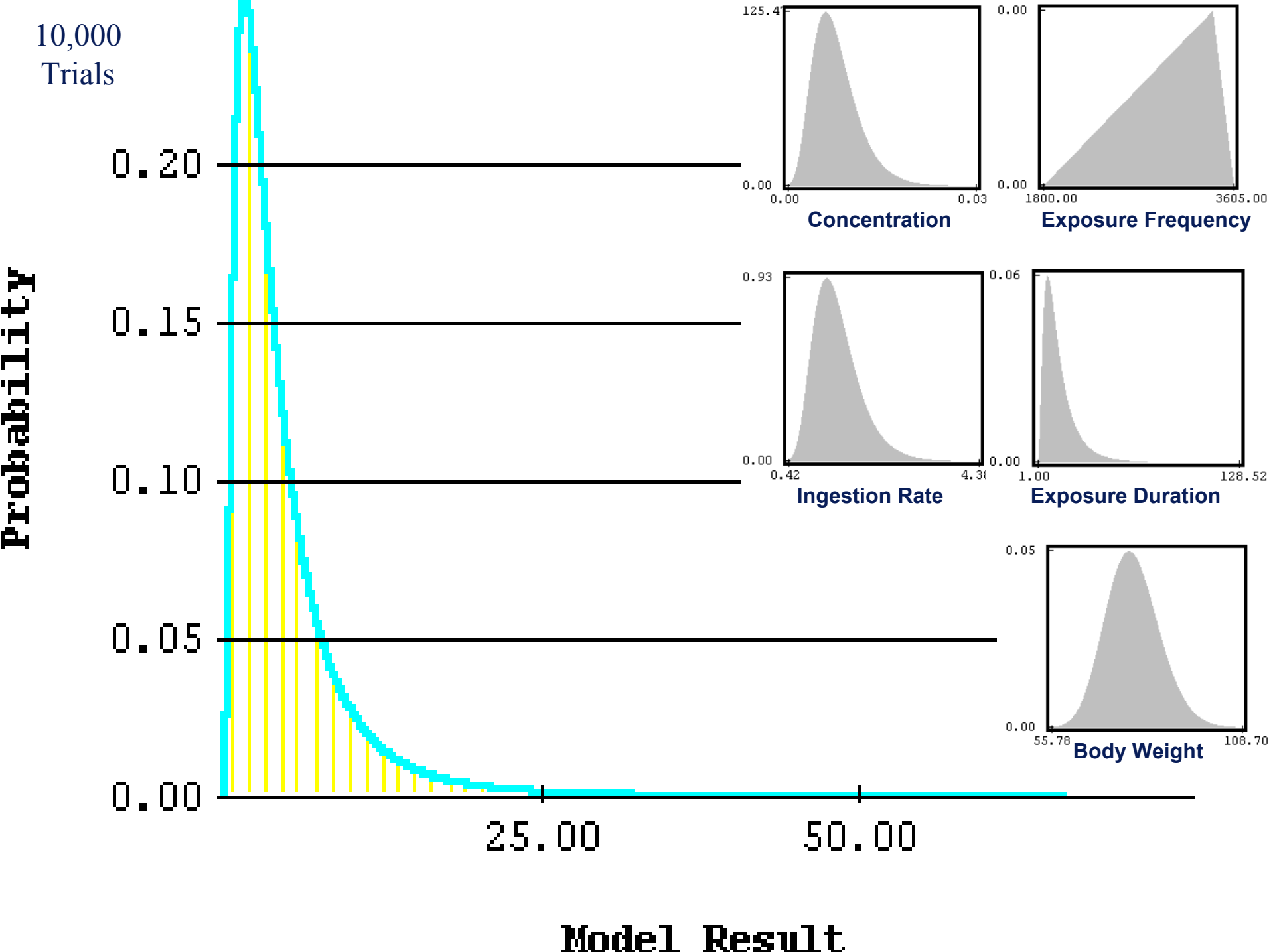
Individual versus Population Risks

Risk Descriptors

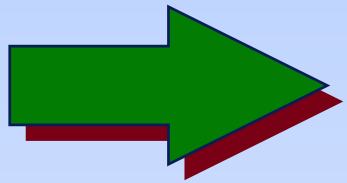
- Central Estimates
- High End
- Reasonable Worst Case
- Theoretical Upper Bound Estimate (TUBE)

Development of Probabilistic Approaches (Monte Carlo)




10,000
Trials
Probability
0.20
0.15
0.10
0.05
0.00
25.00
50.00
Model Result
125.4
0.00
0.00
0.03
Concentration
0.00
0.00
1800.00
3605.00
Exposure Frequency
0.93
0.00
0.42
4.3
Ingestion Rate
0.06
0.00
1.00
128.52
Exposure Duration
0.05
0.00
55.78
108.70
Body Weight

Mechanistic data refines interpretation and extrapolation of :

Dose  Response

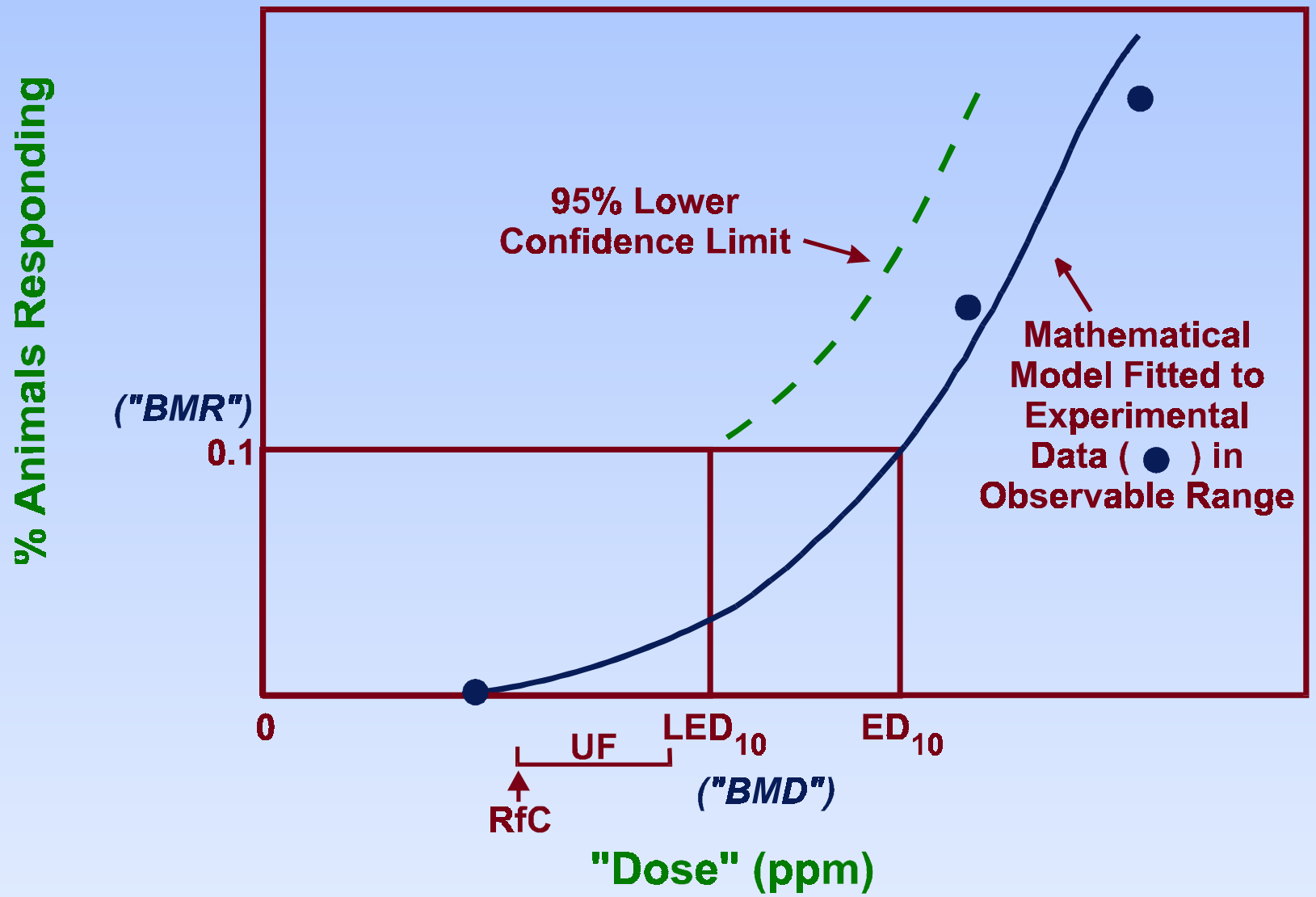
Characteristics of Dose-Response

- Linear
- Sublinear
- Supralinear
- U-Shaped

Comparison of Outputs of Dose Response Analysis

- Probabilistic Estimate of Upper Bound on Risk
- Margin-of-Exposure (M-O-E)
- Reference Dose (RfD)
- Benchmark Dose (BMD)
- NOAEL/LOAEL

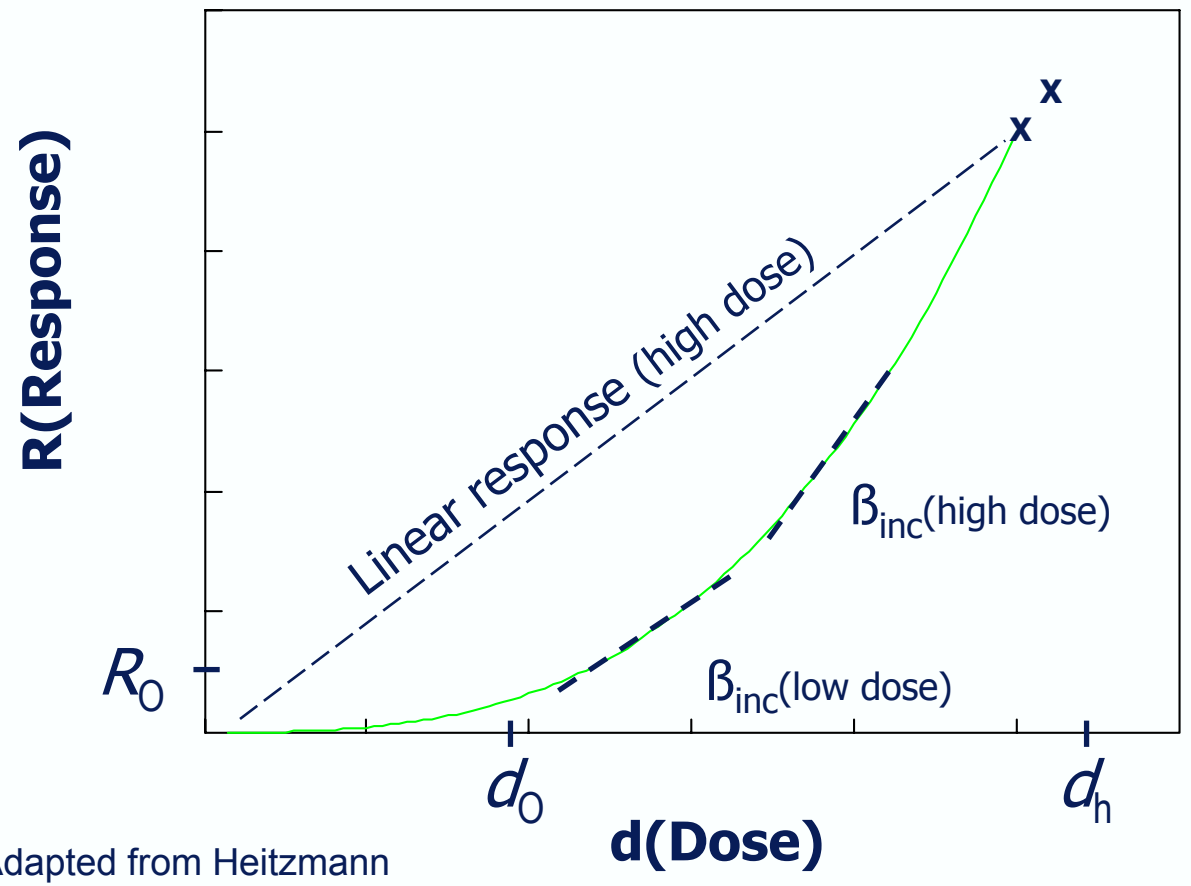
"Benchmark Dose" Approach to Dose Response Analysis for Noncancer Endpoints



Use All Information to Design Cancer Dose Response Assessment

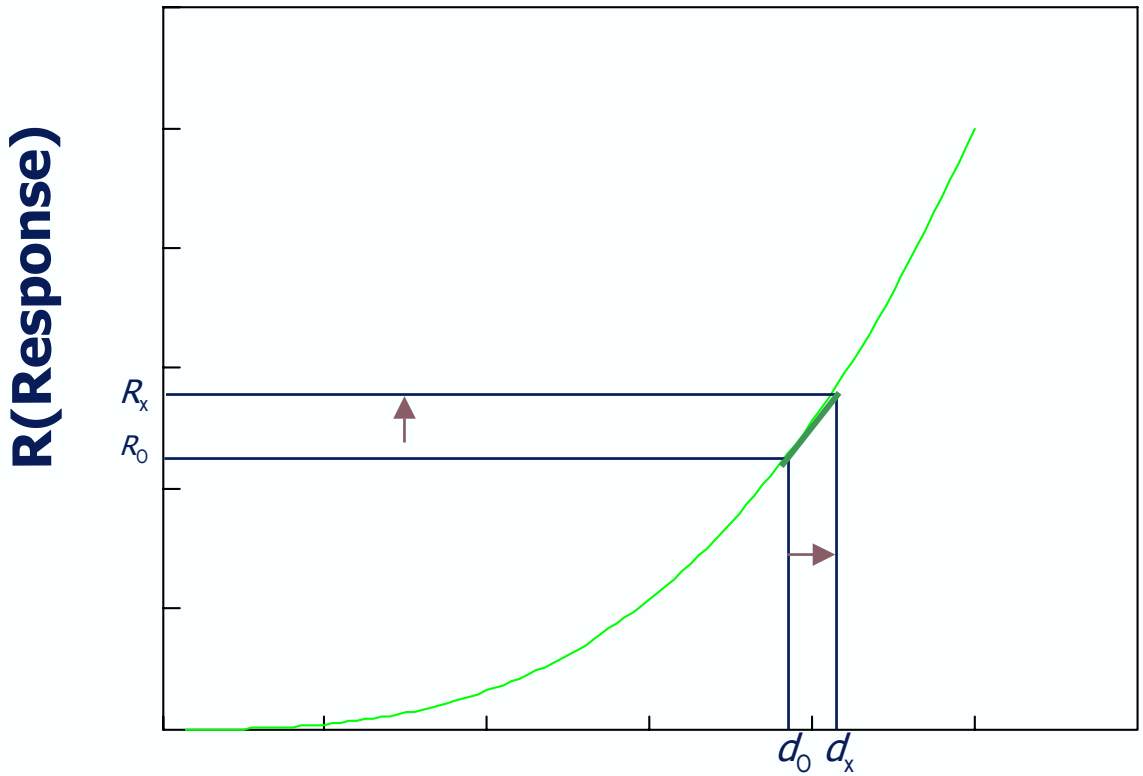
- Tumor data
- Pharmacokinetics and metabolism data
- Data on effects of agent on carcinogenic processes

Comparison of Slopes *



* Adapted from Heitzmann and Wilson (1997)

Additivity to Background *



* Adapted from Heitzmann and Wilson (1997)

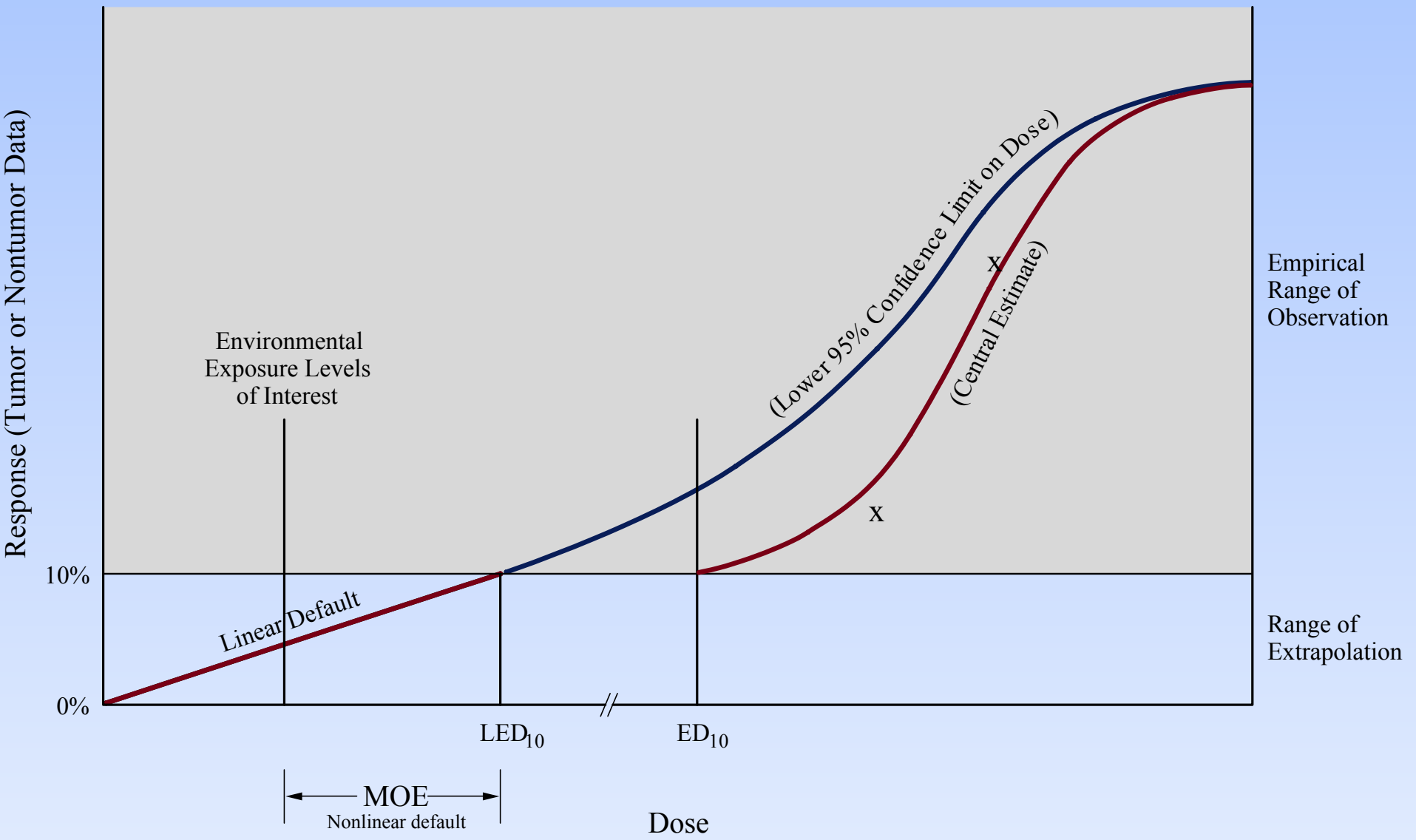
Use of Mode-of-Action Data in Dose Response Assessment

- Construct a biologically-based or case specific model
- Link dose response curve for precursor effect to dose response for tumor effect
- Use dose response for other effect in lieu of that for tumor effect if it is judged to be a better measure of potential risk
- Use to inform assessment of possible dose response in range of extrapolation

Two Step Dose Response Assessment

1. *First step*
   - Data in range of observation
2. *Second Step*
   - Evaluation in range of human exposure (Extrapolation)

Dose Response Assessment



Goal of Probabilistic "best estimate"

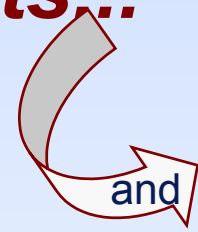
Current Risk Assessment Approaches Raise the Following Issues:

- ⇨ Characterization of subtle, low response biomarkers; protective vs. predictive?
- ⇨ Response biomarkers will be surrogates for effect or multiple effects rather than the effect of concern itself
- ⇨ Additivity to background (exposure, response) may be important to address where exposure of interest lies on the dose-response curve
- ⇨ Outputs are likely to be ranges or distributions

Where do we go from here?

- ☑ Development/validation of sensitive tools aimed at understanding mode-of-action
- ☑ Incorporation of "Framework" Concept
- ☑ More Attention to Route-Specific/ Situation-Specific Characterizations
- ☑ Addressing Sensitive Subpopulations

"Biologically-Based Risk Assessments..."



Biologically-Based Risk Assessment

- Refine estimates of dose to relevant targets through use of biomarkers of exposure
- Improve hazard characterization through use of biomarkers of response with mechanistic linkage to endpoints of concern
- Strengthen inferences regarding the shape of dose/response curves outside the range of observation
- Identify targets of opportunity for further study in potentially sensitive human populations

Appendix E-2

"HAP Benefits Analysis in Section 812 Reports to Congress; Briefing for SAB/EPA Workshop, June 22, 2000,"

Presented by Mr. James DeMocker

HAP Benefits Analysis in §812 Reports to Congress

Briefing for SAB/EPA Workshop

June 22, 2000

Jim DeMocker, EPA/OAR

§812 Benefit-Cost Analyses
Analytical Requirements

- "(a)...The Administrator shall conduct a comprehensive analysis of this Act on the public health, economy, and environment… [which] should consider the costs, benefits and other effects...[of] each standard issued for… (2) a hazardous air pollutant listed under §112, including any technology-based standard and any risk-based standard…"

- "(b)…The Administrator shall assess how benefits are measured in order to assure that damage to human health and the environment is more accurately measured and taken into account."

§812 Benefit-Cost Analyses
Review Requirements

- "(f)...The Administrator shall appoint an Advisory Council… [consisting of] recognized experts in… health and environmental effects of air pollution, economic analysis, environmental sciences, and other [appropriate] fields."

- "(g)…The Council shall review… the data… the methodology… and the findings of such report, and make recommendations to the Administrator concerning the validity and utility of such findings."

§812 Benefit & Cost Estimation

- "Retrospective Study"
  - Submitted to Congress October 1997
  - Direct costs aggregated and fed to macro model
  - Benefits by pollutant as data and models allowed
- "Prospective Study"
  - Submitted to Congress November 1999
  - Direct costs estimated by title / major provision
  - Benefits by pollutant as data and models allowed

Retrospective Study
Stationary Source Pollutants
14 key HAPs from Cancer Risk Study (1990):

- arsenic
- asbestos
- benzene
- 1,3-butadiene
- carbon tetrachloride
- chloroform
- chromium (VI)
- dioxin
- ethylene dichloride
- ethylene dibromide
- formaldehyde
- gasoline vapors
- product of incomplete combustion (PICs)
- vinyl chloride

Retrospective Study
Stationary Source Method

- Incidence change assumed proportional to emissions change

$$I_{ty} = I_{by} \times \frac{A_{ty}}{A_{by}} \times \frac{P_{ty}}{P_{by}} \times \frac{(1 - C_{ty})}{(1 - C_{by})}$$

I = incidence (from CRS)

A = activity (from macro model)

P = population

C = control efficiency (from CTGs, BIDs, regs, experts)

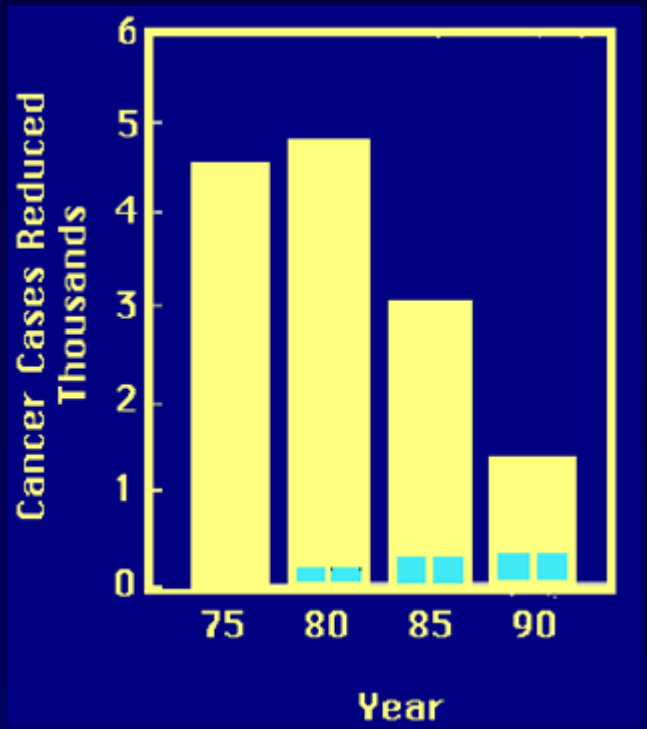
by = base year (85)

ty = target year (70, 75, 80, 90)

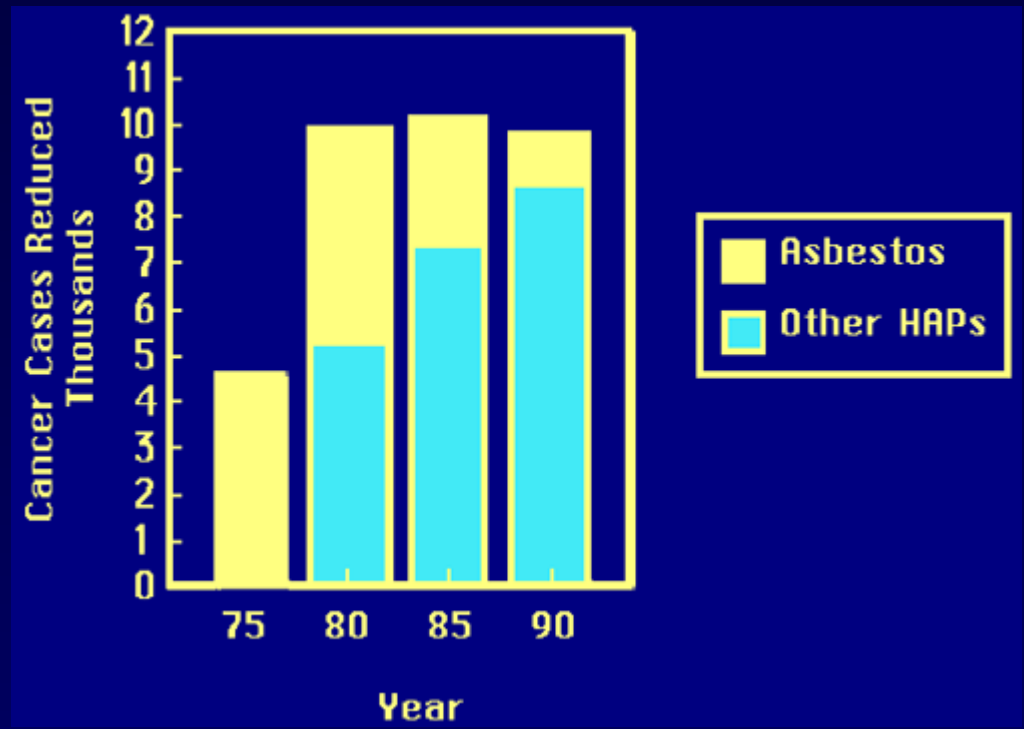
Retrospective Study

Stationary Source Findings

Lower Bound for Other HAPs



Upper Bound for Other HAPs



Retrospective Study
Stationary Source Review Issues

- Estimated incidence for vinyl chloride and asbestos much higher than historical incidence
- Cancer Risk Study designed for only rough order-of-magnitude estimates
  - Unit risk factors are upper-bound estimates
  - Exposure estimates are typically upper-bound (MEI)
- Control efficiencies assumed uniform across facilities and 100% compliance with regulations

Retrospective Study
Report to Congress Presentation

- HAP benefits excluded from primary analysis - described in Appendix
- Quantitative analyses with caveats
  - Stationary source cancer incidence reduction estimates
  - Motor vehicle exposure reduction estimates
- Qualitative discussions
  - non-cancer health effects
  - ecosystem effects

Retrospective Study

Report to Congress

-- Health Research Recommendations --

- Address additional pollutants
- Address mechanisms with pharmacokinetics
- Address variations in human susceptibility
- Address interactive effects of multiple exposures
- Develop alternatives to cancer upper-bound methods
- Develop D/R relationships for non-cancer effects
- Develop methods for acute exposure effects

Retrospective Study

Report to Congress

-- Exposure Research Recommendations --

- Expand data collection: control efficiencies, HAP speciation, facility locations and operating parameters
- Develop more comprehensive exposure models
- Refine uncertainty analysis methods

Retrospective Study

Report to Congress

-- Ecosystem Research Recommendations --

- Estimate levels of bioaccumulating toxics in media
- Correlate levels of bioaccumulating toxics with exposures, concentrations, and adverse effects
- Develop wildlife correlate to RfD or D/R relationship
- Address effects of mixtures
- Address additional ecosystems
- Address wetland species and functions

Retrospective Study
Report to Congress
-- Valuation Research Recommendations --

- Address additional endpoints consistent with kinds of damages expected

- Initiate broad-scope economic valuation using survey techniques

Prospective Study
Methodology Alternatives Presented
-- National Scale --

- Assessment System for Population Exposure Nationwide (ASPEN)
  - Emissions inventory
    - multiple pollutants
  - Air dispersion model
    - point, area, and mobile source categories
  - Exposure model (not completed)

Prospective Study
Methodology Alternatives Presented
-- National Scale --

- Advantages
  - Includes treatment of
    - reactive decay (simplified)
    - secondary formation (simplified)
    - long-range transport (continental scale)
    - wet and dry deposition (parameterized)
  - Emissions/Dispersion well documented:
    - sensitivity analysis
    - model performance evaluation
    - uncertainty analyses

Prospective Study
Methodology Alternatives Presented
-- National Scale --

- Limitations
  - National emission inventory uncertainties
  - Gaussian model limitations
  - Meso-scale transport not addressed (50 - 200 km)
  - Re-suspension not addressed
  - Not stochastic
  - Spatial and temporal peaks not addressed
  - Indoor sources not addressed
  - Indirect exposures not addressed

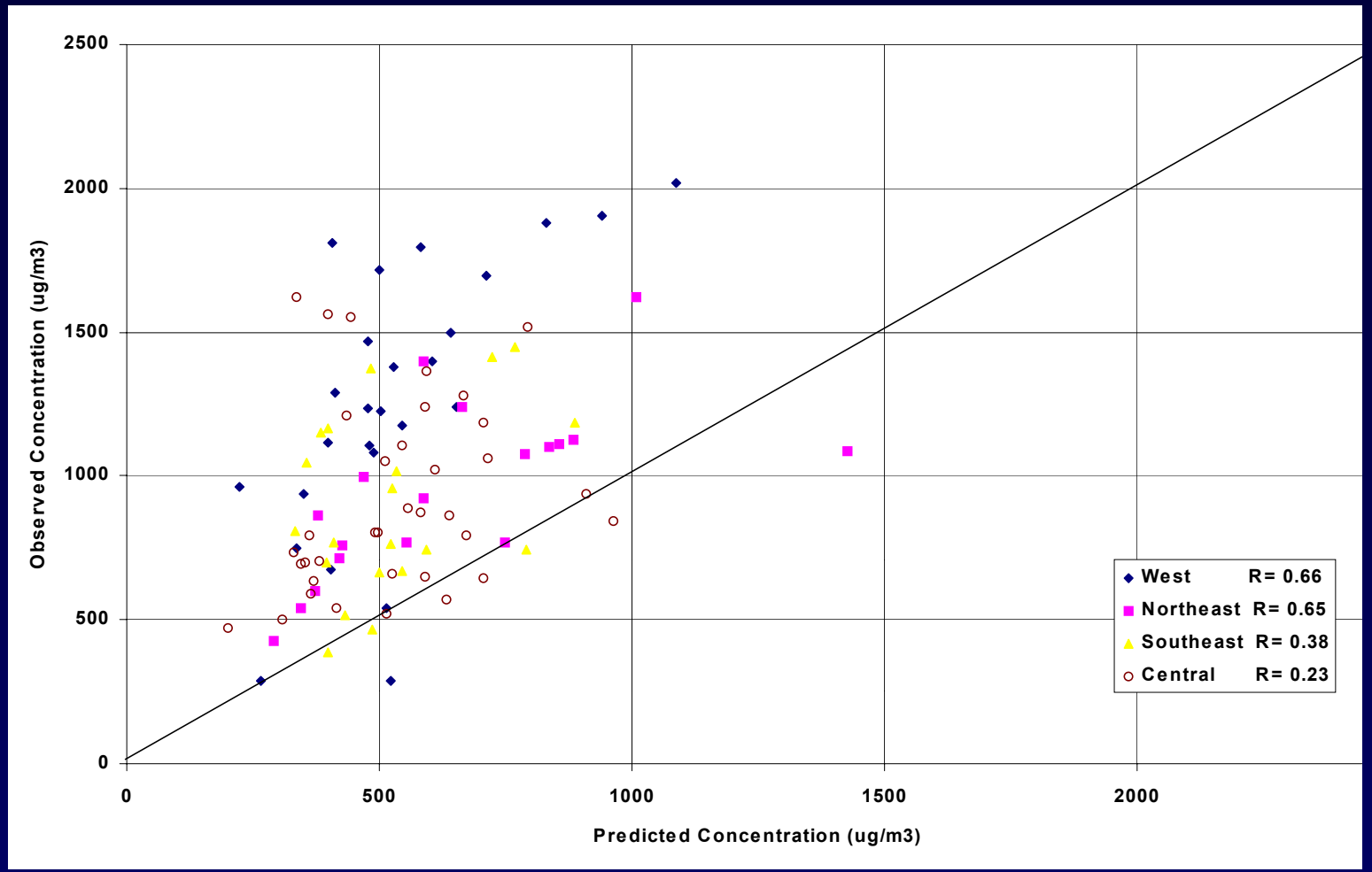
Prospective Study
Methodology Alternatives Presented
-- Local Scale / Case Study --

- Air Quality Integrated Management System (AIMS)
- Developed for Baltimore and planned for Houston and Chicago
- Integrates routinely collected data (measured air quality, emissions, and meteorological data) and dispersion modeling

Prospective Study
Review Issues Raised

- Resources for in-depth analysis for 188 HAPs prohibitive: find priority HAPs
- Unit Risk Factors are upper-bound estimates
- Limited ambient monitoring data to validate ambient concentration estimates
- Exposure assessment limitations
  - 50 km downwind distance for dispersion
  - lack of attention to indirect pathways (e.g., Hg, dioxin)
  - ASPEN preliminary performance evaluation concerns

ASPEN Model Performance
1990 Carbon Monoxide



Prospective Study:
Report to Congress Presentation

- No quantified benefits
- Expect benefits from MACT and incidental to criteria pollutant control
- Besides cancer inhalation impacts, other potential benefits include reductions in:
  - Non-cancer health effects
  - Indirect non-inhalation exposure
  - Ecological and welfare effects

Prospective Study
Report to Congress Research Recommendations

- Workshops to address HAP benefits challenges:
  - toxicology/risk assessment
  - exposure assessment
  - economics
- Investigate use of EPA's Air Toxics Data Archive of measurement data from state / local programs
- Explore whether "supersite" monitoring programs can provide HAP ambient concentration data

Future §812 Studies

- Pondering potential scope, objectives, and reference period for "812 III"
- Detailed analytic blueprint to be developed, and HAP Workshop outcomes will be considered
- SAB Council and HEES will be asked to review analytical blueprint prior to initiation of work

Backup slides

Retrospective Study:
Mobile Source Analysis
-- Methods --

- Based on Motor Vehicle Related Air Toxics Study (1993)
- Exposure estimated for CO using measured concentrations and HAPEM-MS
- Exposure to HAPs assumed proportional to emission factors

Retrospective Study:

Mobile Source Analysis

-- Methods and Data--

$$E = ((A \times C) - B) \times M \times S \times \frac{VOC \times HAP}{CO}$$

E = exposure concentration

A = annual average CO ambient concentration (AIRS)

C = CO ambient to CO exposure concentration ratio (HAPEM)

B = CO background concentration (reported measurements)

M = fraction of CO emissions from mobile sources

S = scenario-to-control scenario CO emission factor ratio

VOC = VOC mobile emission factor by, scenario/year

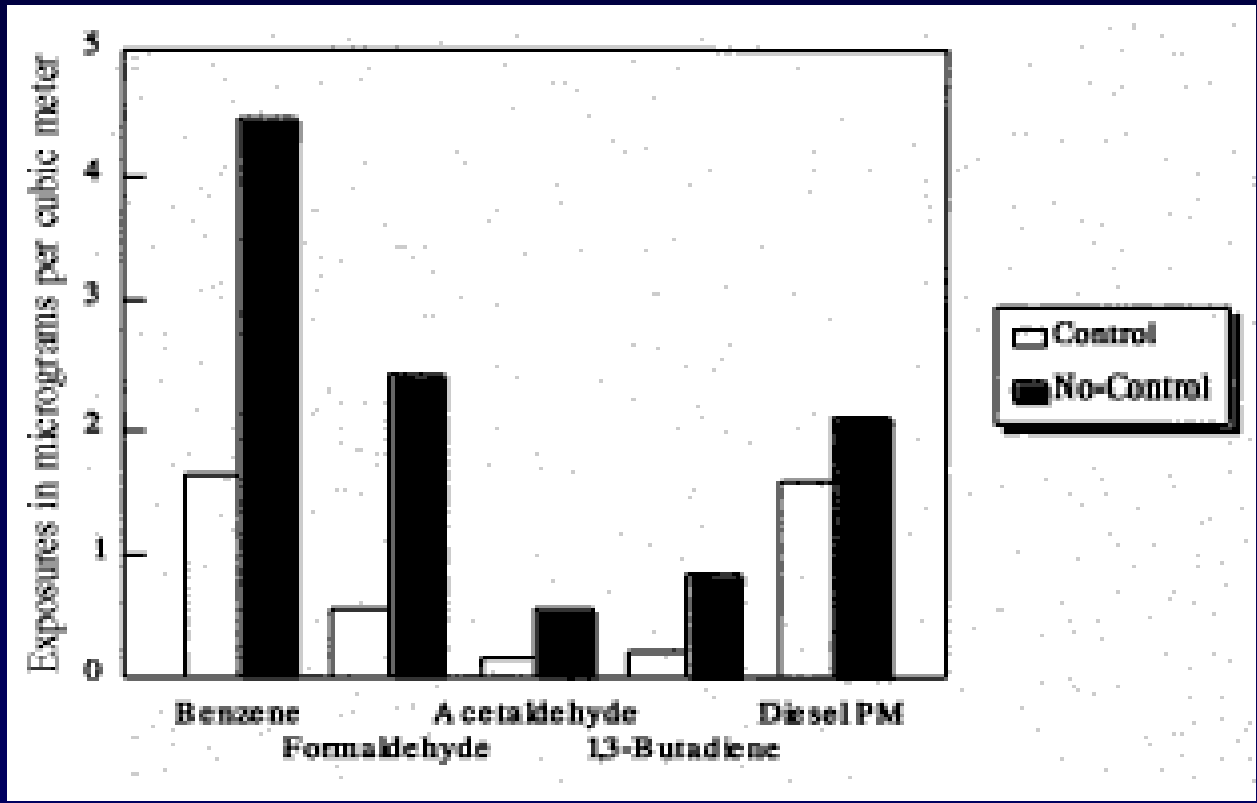
HAP = HAP speciation factor for mobile source VOC, by scenario/year

CO = mobile source emission factor, by scenario/year

Retrospective Study:

Mobile Source Analysis

-- Findings --



Future 812 Studies
Tools Needed

- Expanded air toxics monitoring data
  - 90 new monitors by end of FY00
  - Air Toxics Data Archive to supplement AIRS with state and local data
- Improved emissions inventories
  - 1996 National Toxics Inventory (NTI)
- Evaluation/enhancement of air quality and exposure modeling tools
- Expanded risk data and improved methods

Future 812 Studies
Risk Assessment Workshop

- Current risk assessment state-of-the-art
  - Probabilistic estimates for cancer
  - Reference doses/concentrations for non-cancer
    - More sophisticated D/R assessments for some criteria pollutants
  - Mixtures
    - Sum of upper-bounds for cancer
    - Hazard index for non-cancer

Future 812 Studies
Risk Assessment Workshop

- Recent trends in risk assessment
  - Cancer: mix of probabilistic (no threshold) and reference concentrations (threshold)
  - Non-cancer: modeling and distributional approaches
  - Dosimetry models focused on tissue concentrations

Future 812 Studies
Risk Assessment Workshop

- Potential sources of bias in risk estimates
  - Linear high-to-low dose extrapolation
  - Cross species scaling factor
  - Treatment of untested chemicals and other data gaps
  - Latent effects
  - Use of most sensitive test results
  - Non-cancer uncertainty factors
  - Magnitude and severity of effects
  - Route-to-route extrapolation
  - Benchmark response rate (LED10 instead of NOAEL)
  - Additive treatment of mixtures

Future 812 Studies
Risk Assessment Workshop

- Uncertainty in risk estimates
  - Types
    - Causal link between exposure and effects
    - Magnitude of risk
  - Can use analysis of quantifiable uncertainty to develop central risk estimate
  - Unquantifiable uncertainty may still lead to bias
    - use of sensitive species
    - consideration of non-relevant effects

Future 812 Studies
Risk Assessment Workshop

- Topics for discussion
  - How to characterize a distribution of risk estimates as an input to benefits assessment
  - How to characterize the value of reducing exposure in a reference dose framework: proportion of people above RfD, contingent valuation, other?
  - How to characterize benefits when uncertainty is great: point estimate, range, other?
  - Are some benefits better left unquantified?

APPENDIX F-1

Estimating the Benefits of Abating Toxic Air Pollutants: What do Benefits Assessors Need from Risk Analysis – and Why are they Unlikely to Get it?

White Paper by Dr. Lester Lave,
Graduate School of Industrial Administration,
Carnegie-Mellon University, Pittsburgh, PA.

Estimating the Benefits of Abating Toxic Air Pollutants: What Do Benefits Assessors Need From Risk Analysis – and Why are They Unlikely to Get It?

Lester B. Lave 5-22-00
Carnegie Mellon University

Abstract

I examine what benefits assessors require to estimates of benefits and costs of abating hazardous air pollutants, why current toxicological information is not suited to benefits estimation, and a research agenda that would lead to confident estimating of the benefits of abatement. While it has many limitations and uncertainties, epidemiology data provides information for benefits assessment. In contrast, regulatory toxicology is focused on protecting people from harm, not on giving best (unbiased) estimates of the harm to people from various exposures. Few toxicology tests have been validated in the sense showing that they predict human toxicity. Other than reasoning from first principles, we have no way of knowing which toxicology tests are best are predicting human toxicity and, if so, how much. Fundamental changes in toxicology are required to get data for benefits estimation. In my judgment, the pressure for estimating the benefits and costs of abating hazardous air pollution is putting the right sort of pressure on the regulatory environment. These pressures will force changes in toxicology that will improve identification and quantification of human toxicity.

Introduction

My first task is to inform risk analysts what economists need to estimate the benefits and costs of abating hazardous air pollutants. That is a straightforward task. My second task is to tell economists why the current toxicology information that they are getting is not capable of estimating these benefits. My third task is to open a dialogue between risk assessors and benefits estimators that might some day result in confident estimates of the benefits. My final task is to suggest a research agenda leading to confident estimates of the benefits of abating hazardous air pollutions. To illustrate the first task, I begin with a review of a large EPA benefit-cost analysis.

Assessing the Benefits and Costs of the Clean Air Act

Assessing the benefits and costs of an environmental program, such as the retrospective and prospective effects of the 1970 Clean Air Act, requires information beyond what is factually available. Estimating the costs of the act requires estimating the private and social costs, including abatement costs and opportunity costs, of the regulations. This cost must be contrasted with the costs that would have been incurred if the Clean Air Act and associated regulations did not exist. The actual costs incurred are difficult to estimate; the costs that would have been incurred absent the act cannot be observed or even estimated with confidence, since the world of no regulation does not exist. The cost estimation is, by far, the easier part of the benefit-cost analysis.

The first step in assessing the benefits is to estimate the improvement in air quality

resulting from the Clean Air Act. Unfortunately, there are not good inventories of the ambient air quality or the amounts of each pollutant emitted even today, much less the ambient air quality and amount emitted in 1969. Since the economy has grown since 1970, we need to estimate ambient air quality and how much emissions would occur at each date, with and without the Clean Air Act. This task is similar to that for estimating the costs of abatement.

The second step in estimating benefits is estimating the social benefits that result from lower air pollution levels. The benefits include human morbidity and mortality, visibility, odors, aesthetics more generally, damage to ornamental plants and crops, and other effects from controlling air pollutants. In particular, the following steps are necessary:

Identify each relevant category of harm – eliminate those that are "trivially" small.
Quantify the relationship between ambient air quality and each effect.
For each year, assess the changes in ambient air quality as a result of the regulations.
For each year, estimate the physical benefits of cleaner air in terms of the categories identified in 1 and quantified in 2, e.g., premature deaths, cases of each disease, quality adjusted life-years or disability-adjusted life-years, better visibility, etc.
Value these estimated physical benefits in dollars or compute the value of each relative to the others and compute a benefits index.
Compute the net benefit (benefits minus costs) if the benefits are valued in dollars or the cost-effectiveness (dollar costs divided by the benefits index) if only a benefits index can be computed.

Many non-economists think that 5 is the hardest step, the one that cannot be done rigorously. That is wrong. Methods have been developed to estimate these values. Although there are important uncertainties associated with each method, valuation is not the main source of uncertainty in estimating benefits.

Step 3 adds more uncertainty than step 5, in general. We don't know how the exposure of the average American has changed as a result of the CAA. We have measurements of the criteria air pollutants in a few hundred locations, but translating this information into what people breath when they are outside is difficult. Still more difficult is estimating the amount they breathe during the day, since they spend only a small amount of time outside of buildings or motor vehicles. Additional problems are the health interactions of pollutants, sensitized individuals, and local high pollution concentrations.

Estimating Effects on Humans: Using Epidemiology and Toxicology Information

But the greatest uncertainty is contributed by Step 2. Epidemiology studies the morbidity and mortality in humans that result from exposure to toxicants. Unfortunately, there are a host of problems with the usual epidemiology study. The data are almost always observational rather than experimental. Epidemiologists are tempted to search for associations that were not hypothesized before the study began. The data sets always have important uncontrolled factors influencing the results. The data are imprecise or incomplete because of people lost to follow-up or misdiagnosis. The dose (exposure) is

almost never known with prevision, there are interactions with other environmental exposures and genetic defects, and the full extent of reactions among people who are exposed is unknown. Subtle reactions are almost impossible to identify. For example, some epidemiologists use a relative risk of two as the criteria for having confidence in an observed relationship.

Connoisseurs of epidemiology bemoan the misleading studies and quantification difficulties. More than one has told me that they would prefer to rely on some other approach. That statement reminds me of the story of someone seeking to hire a personal assistant. With two persons to choose between, he interviewed the first and immediately hired the second.

Despite these formidable difficulties, epidemiology observes morbidity and mortality in humans. Toxicology generally studies the effects of a toxicant on a laboratory animal or a cell culture. Occasionally, human volunteers to are exposed to the toxicant at levels that are believed too low to harm the individual. For ethical reasons, the studies with human volunteers keep the dose below the level that would be expected to cause even a small adverse effect. Thus, there is no information on disease; one must extrapolate from an observed effect (not adverse) from an acute dose to disease at a higher dose that is usually present for a long period of time. Despite these formidable difficulties, studies on humans obviate the need to extrapolate from rodents or from cell cultures. These studies can provide precise measurement of the effects of the toxicant in these setting, but say nothing directly about what will be the effect on human morbidity and mortality from relevant exposures.

Problems such as lack of standardization of protocols are small compared to the central problem of toxicology: In only a few cases have the results of in vivo or in vitro studies been compared with human outcomes for the test substance. It is hard to know how to interpret the results of a toxicology study. A well-done study will have high "internal" validity, meaning that it can be replicated within this lab and even in other labs. However, the study has little or no "external" validity in the sense of knowing its implications for human health.

Regulatory toxicology seeks to protect humans against harmful exposure to toxicants. Since it is difficult to draw inferences about the harm to humans from experiments with rodents or cell cultures or to extrapolate from high doses to low doses, without knowing the physiological mechanisms by which exposure to a toxicant causes disease, toxicologists have made a set of "conservative" assumptions that are designed to give a "plausible upper bound" to human toxicity. For example, for cancer the standard assumption is to construct the 95% upper confidence level for the exposed rodents with the steepest dose-response relationship. Other assumptions are made about the shape of the dose-response relationship and the exposure of the population that intended to make sure that the risk to humans is not underestimated.

Hormesis

Another major issue is the large body of data showing that low level exposures to toxicants may improve, rather than harm, health. The usual assumption for cancer risk assessment is that even a single molecule of a carcinogen has a small chance of causing cancer. A great deal of laboratory data on animals and even some data on humans suggests that a tiny dose might improve health.

Hormesis becomes a dominant issue for policy in cases, such as benzene, where most Americans are exposed to benzene at parts per billion concentrations. Most of the estimated cases of leukemia from benzene exposure occur at a few parts per billion. Hormesis suggests that concentrations at this level might improve health, rather than cause leukemia.

Cancer Risk Assessment

Figure 1 illustrates a problem in interpreting toxicology data. These data summarize the outcome of National Toxicology Program lifetime rodent cancer bioassay results for about 1,000 chemicals. A first way of looking at the data is the concordance between rats and mice: 70%. Thus, while there is general agreement between rats and mice, the agreement is far from perfect. If we flipped a coin in order to predict the rat outcomes, we would have a concordance of 50%.

Figure 1

		Rats	
	Carcinogen?	Yes	No
Mice: Carcinogen?			
	Yes	35 (TP)	15 (FP)
	No	15 (FN)	35 (TN)

Another view of the data is the ability of a mouse test to predict the carcinogenicity of each chemical for rats. Viewed that way, when the bioassay is positive on mice and is also positive for rats, that is a "true positive" or TP. Unfortunately, for some chemicals the positive result for mice is negative for rats. This "error" is a "false positive" or FP. When the bioassay is negative in mice and negative in rats, that is a "true negative" or TN. Unfortunately, some chemicals are negative in mice but positive in rats, a "false negative" or FN. The figure shows that, for chemicals that are positive in mice, 70% of them are positive in rats (TP) and 30% are negative in rats (FP). Of chemicals that are negative in mice, 70% are negative in rats (TN) and 30% are positive in rats (FN). The usual interpretation of the NTP is that a chemical that is positive in either rodent species is considered to be a possible human carcinogen. Thus, 65% of chemicals tested are considered to be possible human carcinogens.

These NTP bioassay results are used to classify chemicals as likely human carcinogens. Given the 70% concordance between rats and mice, what is the likely concordance between rodents and humans? Since rats are more similar to mice than either are to

humans, it seems likely that the concordance between rodents and humans will be less than 70%. What proportion of the 65% of chemicals that are positive in rodent tests are human carcinogens? How many false positives and false negatives are there likely to be in this classification? As noted above, we know that there are some false positives (the alpha-2-u globulin chemicals) and some false negatives (benzene). The NTP interpretation is intended to minimize the number of false negatives, even at the cost of additional false positives.

It seems unlikely that the concordance between rodents and humans is as high as 70%. In fact, it is more than just possible that the concordance between rodents and humans is no better than 50%, equivalent to a coin flip, which is a bit cheaper than spending more than $1 million on a lifetime rodent bioassay.

This is relevant because of the limited data on toxicity in humans. For example, there is epidemiology data connecting each chemical to a cancer for about two dozen chemicals or groups of chemicals. Unfortunately, the concentrations that people were exposed to are often highly uncertain and so there is only a remote idea of the dose-response relationship. For the other 600 plus toxicants in the Toxic Release Inventory report, there is at best toxicology data, generally in the form of rodent studies. Biologists warn that extrapolating between species is perilous. Using toxicological data to estimate the risks to humans from exposure to a toxicant does precisely that extrapolation. The extrapolation begins with a leap of faith that the effect observed in rodents or in cultured cells predicts human toxicity. The differences in anatomy and physiology between humans and rodents means that many diseases/conditions are unique to rodents or to humans. Still more uncertain is extrapolating human risk from the dose-response relationship observed in rodents.

Non-cancer Risk Analysis

Non-cancer risk analysis proceeds by establishing the no observable effects level (NOEL) in rodents and extrapolating to humans by using safety factors to account for differences among species and for the most sensitive individuals. The practice assumes that humans are the most sensitive species, despite considerable data showing that other species are often more sensitive, e.g., dioxin in mice.

Several proposals have been published to estimate harm for exposure above the reference dose or to make use of a combination of data and judgment. However, in none of this work is there an attempt to give an unbiased estimate of the effect on humans. Past data suggest that for both cancer and other health effects rodents sometimes are positive when humans are negative and vice versa. Furthermore, when both are positive, sometimes the rodents are more sensitive than humans and vice versa. These data suggest that if one knew no more than the results of an in vivo experiment, one would be very uncertain about the implications for humans, both in terms of whether the chemical is a human toxicant and, if so, what is the potency.

I don't mean to suggest that benefits estimators are prissy, unused to highly uncertain data. Indeed, anyone who has read the benefits assessment from the 812 report might be

inclined to ask, as an epidemiologist did of me some years ago: "Are there any data so bad that an economist would not analyze them?" However, the economist needs unbiased estimates, even if they are highly uncertain. Toxicologists are not providing unbiased estimates - that is the fatal flaw.

A Research Agenda

Risk analysis could be most helpful to benefits assessment if toxicologists performed the following analysis:

For cancer, for all the known human carcinogens, examine whether a standard National Toxicology Program bioassay would be positive. If there is some uncertainty, what is the likelihood that a particular bioassay would be positive?

Compare the estimated dose-response relationship for rodents with the relationship for humans, accounting for the uncertainty in the resulting estimates.

1. For chemicals that are positive in the NTP bioassay, which are not or are unlikely to be human carcinogens?
2. For non-cancer endpoints, the same questions are relevant, although there is no single standard for in vivo or in vitro tests. In particular, what is the concordance between different species in the same test and across different in vivo and in vitro tests? What is the concordance between test outcomes and human toxicity data on each chemical?

The other issue is extrapolating from high human doses, observed in accidents, occupational exposure, or tests with human volunteers, as well as extrapolating from high doses in in vivo and in vitro tests to the low doses over long periods that most people experience. In the absence of knowing the mechanism of action, one must rely on assumptions about the nature of the dose-response relationship. The ED01 experiment attempted to pin down the best dose-response relationship for cancer. Unfortunately, all of plausible models performed about as well in explaining the observed data, even though they had very difference implications for the effects at low exposures.

Summary and Conclusions

As currently practiced, regulatory toxicology cannot provide data to estimate the benefits of abating hazardous air pollutants. No minor patches will provide these data. Regulatory toxicology is built on a foundation of protecting humans from harmful exposure to toxicants. It is saturated with implicit and explicit assumptions that lead to "plausible upper bounds" rather than "best estimates" of the exposure-response relationship.

If we knew the mechanisms of action by which toxicants harm humans, we could establish standards that would protect them and also provide best estimates of the exposure-response relationships to the benefits assessors. At present, few mechanisms of action are known and it seems doubtful that we will ever know the mechanisms of action for most toxicants.

Short of knowing the mechanisms, research can do much to clarify the qualitative and

quantitative risks to humans of exposure. Toxicologists need to analyze the implications of current in vivo and in vitro tests for human toxicity. They need to look for human data to compare with laboratory results. I have no doubt that when this happens, we will find that some popular tests predict human toxicity no better than flipping a coin. For these tests, society is wasting its resources and using meaningless data to make regulatory decisions. Other tests can be modified to increase their human predictivity. New tests can be developed that are more predictive of humans.

Insisting on estimating the benefits of reducing exposure to hazardous air pollutants will lead to better policy. More important, it should trigger a revolution in toxicology in searching for laboratory tests that are more predictive of human toxicity.

Appendix F-2

Benzene White Paper

White Paper by Dr. Bernard Goldstein
Environmental and Occupational Health Sciences Institute, Robert Wood Johnson School of Medicine, Rutgers, Piscataway, NJ.

Benzene White Paper

Bernard D. Goldstein, M.D.
Environmental and Occupational Health Sciences Institute
170 Frelinghuysen Road
Piscataway, NJ 08854
Telephone:732-445-0200

June 16, 2000

Abstract

Benzene is a known cause of aplastic anemia and of human leukemia. At community air pollution levels on which the benefits analysis for benzene control are to be based, there is no firm evidence to support a non-neoplastic effect. Estimation of the leukemic effect at such levels requires extrapolation across about three orders of magnitude of benzene dose. There is currently insufficient evidence to depart in any direction from low dose linearity.

Introduction

Benzene has been chosen as one of three compounds to be Case Studies for the **SAB/EPA Workshop on the Benefits of Reductions in Hazardous Air Pollutants: Developing Best Estimates of Dose-Response Functions.** The goal of the Workshop is to discuss dose response assessment methods for hazardous air pollutants (HAP) that are useful for assessing the benefits of emission control measures. This document is intended to provide a background for this discussion.

The literature on benzene toxicity is perhaps as large as that for any of the compounds designated as HAPs under the US Clean Air Act. This information has been extensively reviewed elsewhere (Caprino and Togna, 1998; Goldstein and Witz, 2000; Snyder et al, 1993; Smith and Fanning, 1997; Goldstein, 1977; Benzene '95 Conference, 1996; Krewski and Snyder, 2000). I will focus on those studies that may be particularly useful for providing the information needed for economic analysis of the benefits of reducing benzene exposure specifically related to the control of HAPs under the 1990 CAA Amendments. I have been asked to do this relatively late in the process. There has not been time to go through a detailed analysis of the basis for the different risk assessments for benzene, nor do I have the requisite expertise to clearly explicate the major differences in mathematical modeling approaches. Perhaps this is an advantage.

The organizers of the workshop have commissioned an "Economist's Perspective" by Lester Lave, a noted economist who has made major contributions to the economic analysis of air pollution health effects. This is not the place to respond to all of the issues raised by his very provocative piece which unfortunately demonstrates how poorly regulatory biological scientists have communicated with economists, even ones active in the field of risk assessment such as Dr Lave. However, it does lead to a recommendation.

To better understand the interface between regulatory risk assessment and economic benefit analysis, we should rephrase the benzene-related question being asked. The current question is how to estimate the benefit of partially reducing community benzene exposure now in the range of a few parts per billion. Instead, I suggest that EPA develop a hypothetical example of an economic analysis aimed at determining the benefits of reducing a putative workplace standard for benzene of 30 ppm, TWA to a level of perhaps 15 ppm, TWA. Thirty ppm is a useful baseline because it is in this range that there are data about actual leukemia risk and on non-cancer endpoints as well. A target of 15ppm is suggested because it represents only a two-fold reduction to a level that is still within the range of existing human and animal data. A two-fold reduction in outdoor benzene levels is also presumably a reasonable outcome of the new MACT strategy for benzene. The goal of this

exercise would be to obtain a better understanding of the biological uncertainties in the rich benzene data base which impact on a benefit analysis when there is no need for extrapolation to much lower exposure levels. Once this has been clarified, we can more readily address the extrapolation issues relevant to the perhaps three orders of magnitude lower levels of benzene exposure needed for HAPs benefit analysis.

An additional point about the history of HAP regulation relates to this Workshop. Two major driving forces for the 1990 CAA amended approach to regulate HAPs were impatience at the previous rate of EPA's regulatory approach and the impact of TRI data showing the many tons of unregulated pollutants being released into the air . Using a process which required an initial finding of likelihood of adverse effects, relatively few agents were previously regulated under Section 112 of the CAA. Benzene was one of them and in fact there is a clear statement of the use of risk assessment and of economic analysis in the 1984 benzene decision document (EPA, 1984). This included a table describing the costs and the number of leukemia deaths averted for each of the control approaches that were considered (Goldstein, 1985). Further, as only certain of these controls were then imposed, one could relatively easily reconstruct the risk benefit criteria underlying the decision.

In contrast, the present CAA lists more than 180 compounds to be regulated by EPA, in essence shifting the regulatory burden from the government which had to make an initial finding of likely adverse effects before listing, to industry who now must bring sufficient evidence that there are no adverse effects in order to delist. Congress moved away from risk assessment as the primary basis for regulation to a technology based approach in which risk considerations only come into play after Maximum Available Control Technology (MACT) has been instituted. For almost none of these previously unlisted compounds is there the rich data base available for benzene or for the other compounds previously chosen for listing under Section 112. Further, for many of the compounds for which there are ample human data, such as the alkyl benzenes, the data strongly suggest that no measurable adverse effects are likely at community exposure levels. Yet Congress clearly intended such compounds to be subject to MACT control irrespective of the lack of convincing data showing adverse effects. It is inconsistent and perhaps disingenuous of Congress to both disavow risk assessment as the primary basis for the regulatory control of HAPs while at the same time insisting that EPA use risk assessment as a means of demonstrating the benefit of the technology-based regulatory approach which it has imposed.

Health Effects of Benzene

Benzene has been known to produce destruction of the human bone marrow since the 19th Century. In humans, and in all laboratory animals tested, benzene produces a dose-dependent destruction of bone marrow precursor cells that are responsible for the production of mature red blood cells, platelets, and granulocytic and lymphocytic white blood cells. This is accompanied by chromosomal damage. The result is a decrease in all formed elements in the blood known as pancytopenia. A severe form of pancytopenia, aplastic anemia, includes marked loss of bone marrow cellularity and is a frequently fatal disorder.

Benzene is also a known human carcinogen, indisputably causing acute myelogenous leukemia and its variants (collectively called Acute Nonlymphocytic Leukemia - ANLL). It is also likely to cause other hematological tumors. An intermediate diagnosis between benzene induced pancytopenia and ANLL is myelodysplasia, a preleukemic condition characterized by morphological abnormalities and an increase in number of bone marrow precursors representing a monoclonal expansion. Both myelodysplasia and ANLL can occur without being preceded by clinically overt pancytopenia.

At concentrations well over 100 ppm (320 mg/m^3) benzene also causes central nervous system anesthetic-like effects common to alkyl benzenes and other VOCs. While an occupational hazard in enclosed spaces, this non-hematological effect is clearly not pertinent to considerations of HAP control and will not be discussed further. Based on relatively weak evidence in animal studies, more information on the potential for benzene-induced developmental effects in humans would be welcome, but again there seems to be no basis for considering such effects in the present document.

Benzene Exposure Levels Pertinent to Economic Analysis of the Impact of Control of HAPS

In order to provide a discussion of benzene toxicity pertinent to the purposes of this Workshop I have briefly and incompletely reviewed the literature concerning expected outdoor community concentrations of benzene. The focus is on the question of whether non-cancer hematological effects might occur. Accordingly, the goal is to pick a level that would represent a high community exposure that would be a reasonable target for assessing the health and economic impact of MACT control measures. For neoplastic endpoints, if one assumes a linear risk then the absolute level is not important – only the extent of reduction in benzene exposure and the size of the population is needed to calculate the number of leukemia cases averted. But for non-cancer effects the presumption of a no-effect level requires that some attempt be made to determine likely high community exposure levels.

Community outdoor benzene levels even in reasonably polluted areas appear to range well below 10 ppb (32 ：g/m^3). Wallace and his colleagues have measured outdoor air benzene concentrations in various parts of the United States as part of the Total Exposure Assessment Measurement (TEAM) study. In an overview of these studies the mean outdoor air benzene concentration based on backyard measurements of 175 homes in six urban areas was 6 ug/m^3 (Wallace, 1991). The highest levels in the TEAM study came from one of the two studies in Los Angeles where the outdoor air concentrations appeared to range up to 30 ug/m^3 with a geometric mean of 16 ug/m^3 (Wallace, 1986). In a relatively polluted area of New Jersey, the mean levels were 4.1 ug/m^3 at night and 3.8 ug/m^3 during the day. Lagrone (1989) reported outdoor benzene levels in a network of six sites located in an industrial area of Houston ranged from 1.4-5.8 ppb, mean 3.6 ppb, (4.5 - 18.6：g/m^3, mean 11.5 ：g/m^3) during the period September 1987 to March 1988. Johnson et al (1991) used a variety of models to estimate incremental ambient benzene concentrations to receptors living near seven bulk gasoline storage facilities in North Carolina. The highest modeled fenceline annual average benzene incremental concentration was 2.1 ppb (6.7：g/m^3). EPA reported that 1991 benzene concentrations in Lima, Ohio, ranged from 1.1 to 6.8 ug/m^3, mean 2.6 ug/m^3. EPA has also concluded that the background concentration of benzene, attributable to long range transport and non-anthropogenic sources, was 0.48 ug/m^3 (Woodruff et al, 1998). Data from 97 samples taken as

part of the NHEXAS study in EPA Region 5 shows a median benzene level of 2.9 ug/m^3 and a 90th percentile level of 5.6 ug/m^3 (Clayton et al, 1999). There appears to be good evidence that ambient benzene concentrations are decreasing. Based upon monitoring network data, the California Air Resources Board (CARB, 1997) has estimated a population-weighted annual concentration of 14.7 ug/m^3 benzene in 1982 and 2.3 ug/m^3 in 1996.

Benzene Toxicology

To a toxicologist benzene is both fascinating and frustrating. It is a well studied compound that has provided much insight into general toxicological mechanisms of action and is particularly relevant to understanding target organ toxicity related to the bone marrow. Its metabolism has also been thoroughly evaluated and, although complex, is reasonably well understood (Snyder and Hedli, 1996). We also know that it is one or more benzene metabolites, not benzene itself, that is responsible for its hematological toxicity. Yet the linkage between benzene metabolism and benzene hematotoxicity remains elusive. It is not at this time even certain that the toxicological mechanisms by which benzene destroys bone marrow precursor cells leading to aplastic anemia are the same mechanisms producing cancer of these cells. What we do know suggests that Occam's Razor is dull, that there is not a single benzene metabolite producing a single mechanism of cell damage and eventual mutation or cell death (Goldstein, 1990). Rather there are multiple metabolites producing effects in multiple biological pathways that lead through a variety of mechanisms to adverse effects (Chen and Eastmond, 1995B; Eastmond et al, 1987; Guy et al, 1991; Levay, 1992).

Understanding the relation between benzene metabolism and its mechanism(s) of toxicity is one part of the puzzle that must be solved if we are to move away from the routine default assumption that leads to linear extrapolation from high to low dose. A second part of this puzzle is to understand the relation between the observed biological effects of benzene metabolites and the mechanism(s) of carcinogenesis. The available information on these two parts of the puzzle does not always point in the same direction. For example, there is evidence suggesting that the metabolism of benzene to active intermediates saturates at higher doses which could mean that the dose response is supralinear. There is also evidence suggestive of aneuploidy being an important mechanism of benzene carcinogenesis, and it has been argued that such gross chromosome damage requires multiple "hits" indicating that the dose response to lower benzene levels is sublinear. There are counter arguments to each of the above.

Risk assessors have attempted to tease the biology out of the epidemiological findings. For example, Crump (1994) using a weighted exposure approach has calculated the apparent latency period for benzene leukemogenesis from the pliofilm cohort database and has inferred possible biological explanations for why his calculations result in a longer latency period for leukemia than was observed in radiation-exposed atom bomb survivors. Inferring biology from epidemiology can be useful for hypothesis generation but needs to be approached cautiously as a basis for risk assessment.

Current advances in molecular biology, including studies using these techniques in benzene-exposed humans and animals, provide much promise for the eventual unraveling of the mechanisms of

benzene hematotoxicity and leukemogenesis (Chen and Eastmand, 1995A; Rothman et al, 1995; Xu et al, 1998; Irons, 2000; Laskin, 2000; Mani et al, 1999; Smith and Rothman, 2000). But at the present time it is difficult to observe a clear pattern, or even a clear directional signal, that would permit a generally accepted biological basis for other than a classic linear model of benzene leukemogenesis.

Risk Assessment for Hematological Neoplasms Caused by Benzene.

The most recent EPA update on benzene (EPA, 1997) derives two risk estimates from the pliofilm cohort: a lifetime leukemia risk at 1ppm (3.2 mg/m^3) of 1.8×10^{-2} using an additive risk model, and 4.1×10^{-2} using a relative risk model. Both are based on linear low dose assumptions. These are little changed from previous EPA risk estimates of 2.6×10^{-2}, based primarily on the geometric mean of four maximum likelihood risk estimates (EPA, 1985) or of an even earlier risk estimate of 2.2×10^{-2} risk at 1ppm (3.2 mg/m^3) (Goldstein, 1985). The EPA (1997) NCEA review of different approaches based on the pliofilm data that use linear assumptions states that the risk at 1ppm (3.2 mg/m^3) ranges from 4.7×10^{-3} to 2.5×10^{-2}. Assuming low dose linear extrapolation this translates into a risk of 47-250 in a million for a 70 year lifetime exposure to 10 ppb benzene (3.2 mg/m^3), a reasonable upper bound for a community exposure level. Of note is that there is a reasonable similarity between benzene risk assessments derived from the human and animal cancer data (Goldstein, 1985).

There are three areas of uncertainty that are particularly pertinent to debates concerning the appropriate risk of benzene-induced cancers: (1) the extent of benzene exposure of workers in cohorts with an increased risk of ANLL, particularly the pliofilm cohort; (2) the appropriate shape of the dose-response curve for extrapolating the carcinogenic potential of much lower level benzene exposures; and (3) whether benzene also causes hematological cancers other than ANLL. The major uncertainty is the shape of the dose-response curve and particularly whether there is sufficient evidence to deviate from low-dose linearity.

(1) <u>Extent of benzene exposure in cohorts with an increased risk of ANLL</u>

One of the most thoroughly evaluated cohorts in the history of occupational epidemiology has been that of pliofilm workers in two Goodyear plants in Ohio. The increase in leukemia incidence among these workers put an end to any remaining doubt that benzene was a cause of ANLL (Infante et al, 1977). As benzene exposure levels had been reported within the then allowable 10 ppm TWA workplace standard, OSHA attempted to impose an Emergency Temporary Standard of 1 ppm, which was thrown out by the Federal court. During the formal rule making that followed, it became apparent that the pliofilm workers had in fact been exposed to benzene levels well above the standard. Identifying the actual exposure levels became particularly important to establishing the new workplace standard, and to establishing the leukemogenic risk of benzene. Rinsky and his colleagues at NIOSH performed what was then the most extensive retrospective cohort exposure assessment (Rinsky et al 1981). Not surprisingly, they were forced to make numerous assumptions as to past exposure levels. Particularly controversial was their assumption of what seemed to be relatively low levels of exposure during the World War II period. As pliofilm production was an

essential war industry, and a nationwide occupational health review made during this period indicated a rather cavalier use of benzene in such industries, it seems reasonable that exposure levels were significantly higher. Building on the work of Rinsky et al (1981), Crump and Allen (1984) Paustenbach et al (1992, 1993) and Paxton et al (1994A) performed extensive reanalyses of the exposure levels. Using the reported blood counts in these workers, my colleagues compared the Crump and Allen with the Rinsky analyses and found that the former more closely predicted the blood count variations (Kipen et al, 1988; Cody et al, 1993). But we pointed out that our findings were not relevant to the absolute exposure levels, only to the relative time variations.

The NIOSH researchers have vigorously defended their exposure assessment (Utterback and Rinsky, 1995). A more recent exposure analysis of the pliofilm cohort has been performed by Schnatter et al (1996) focusing on exposure levels of various worker subgroups in order to refine the risk estimates. However, EPA (1997) concluded that the various exposure assessments for the pliofilm cohort do not differ among themselves sufficiently to have a major impact on the 1985 estimate of risk based upon the epidemiologic data alone. Note that the debate about benzene exposure levels also has implications to mechanistic issues concerning dose rate and linearity.

Recent studies in China by scientists from the Chinese Academy of Preventive Medicine and the US National Cancer Institute have provided another data base from which one can attempt to relate an elevated risk of ANNL to workplace benzene exposure levels (Yin et al, 1987A, B, 1996A, B; Zhang et al, 1996; Rothman et al, 1995, 96; Hayes et al, 1997). The number of leukemia deaths is appreciably higher than that for the pliofilm cohort. A reconstructed exposure estimate for much of the cohort has been reported (Dosemici et al, 1994). While a very useful exercise, the inherent uncertainties in this dose reconstruction are already leading to controversy (Wong, 1998, 1999).

Preliminary review suggests that the resulting dose response estimates will be in the similar range of the leukemia dose response estimates for the pliofilm workers. However, it is still possible that ongoing prospective and retrospective studies of these heavily exposed workers will lead to more refined estimates of dose response patterns. Of perhaps greater importance to the issue of low dose linearity is the mechanistic information that may be obtained from study of these benzene-exposed workers.

Understanding the dose portion of the benzene dose-response relationship requires knowledge of the specific workplaces. For most industrial settings there tends to be large variations in the extent of individual exposure which is often not apparent from usual industrial hygiene measurements. Workers may be in a part of the refinery or chemical factory that is particularly prone to have a leaking valve. Or there may be individual habits that should be prevented, such as using benzene to wash off grease and grime from hands or clothes. Smaller and unregulated workplaces, which seem to characterize the Chinese experience, may well have a larger degree of individual variation. This variation is important, particularly for establishing leukemia risk, because fortunately only a small percentage of even a highly exposed workforce develops ANNL. Area benzene measurements may not reflect the actual exposure of those relatively few individuals, raising questions as to the validity of risk assessments based on such measurements, particularly in workplaces with highly variable exposure conditions.

(2) The linearity of the dose response curve for leukemogenesis

Extrapolation of benzene carcinogenesis from the benzene levels observed in the pliofilm or Chinese studies to the three or so orders of magnitude lower levels pertinent to community exposure levels has been highly controversial. There have been numerous analyses that have argued that these data, or others, provide evidence of a non-linear, sublinear or threshold for benzene leukemogenesis (see for example Paxton, et al, 1994B: Cox, 1996; Schnatter et al, 1996; Wong and Raabe, 1995) that would lead to a substantial decrease in the estimated risk for the lower level exposures relevant to community benzene exposures. EPA (1997) has claimed that more than 100 risk estimates have been presented, and that they vary by 6 orders of magnitude at 1 ppb. The key issue is whether the extrapolation is assumed to be linear or non-linear. In essence, EPA has defended linear extrapolation as the preferred approach unless there is adequate biological evidence supporting a different dose response relationship.

The most extensively analyzed cohort of benzene exposed workers to date has been that of the pliofilm workers (Infante et al, 1977; Rinsky et al, 1981, 1987). A key issue is that with only 9 cases of leukemia in the Rinsky, 1987 follow-up, versus 2.66 expected, there is very little stability in any of the analyses. This follow up study may not have been warranted given the relatively short latency period for ANLL as compared to solid tumors and the higher benzene exposures in the past. In essence, additional follow up may dilute out the true effect and only add cases that are unrelated to benzene exposure, thereby obfuscating the dose response issues even further.

Hayes et al (1997) did an extensive analysis of hematologic neoplasms in Chinese workers, reporting on a cohort of 74,828 benzene-exposed and 35,805 unexposed workers. Their key finding was that for workers historically exposed to benzene at levels of 10 ppm (32 mg/m^3) or less there was an elevated relative risk for all hematological tumors combined of 2.2 (95% CI 1.1-4.2). For ANLL and myelodysplasia the RR was 3.2 (95% CI 1.0-10.1). When the exposure levels were consistently 25 ppm or more, the RR for ANLL and myelodysplasia was 7.1 (95% CI 1.1-15.9). Risk for non-Hodgkin's lymphoma, an unusual tumor in China, was also significantly increased. The authors cautiously note that the dose response curve for benzene-induced cancer from their study tended to flatten out suggesting a supralinear curve. The cohort was also reported upon with slightly different numbers (Yin et al, 1996B). This does response estimation has been criticized by Wong (1998, 1999) and by Wong and Raabe (2000).

The Health Council of the Netherlands reviewed benzene risk in 1987 and again in 1997 (Health Council, 1997). In 1987 the conclusion was that benzene was a human carcinogen and that it worked through a genotoxic mechanism. It was considered uncertain as to whether the mechanism of action was stochastic or non-stochastic, i.e. without or with a threshold. A linear extrapolation was chosen as it was deemed to be the more cautious approach. However, the Health Council believed that a simple linear extrapolation was not warranted as it would produce "excessively low" results that would be "overly safe". Accordingly they chose a factor of 100 higher than would otherwise result from a linear extrapolation to a one in one million lifetime risk. This resulted in a recommended exposure limit of 12 ug/m^3 of benzene in outdoor air.

Their more recent evaluation used a circuitous route to confirm the previously recommended level. The Health Council's benzene committee again stated itself as being uncertain as to whether benzene has a stochastic or non-stochastic mechanism of action. However, they were impressed by a study of 208,000 petrochemical employees said to have an average exposure to 0.7 mg/m^3 benzene which they interpret as showing no increase in ANLL (Wong and Raabe, 1995). They extrapolated this as being equivalent to 35ug/m^3 lifetime to the general population. They further concluded that this supported their earlier view that the exposure-response curve will be sublinear rather than linear. As there was still uncertainty as to the exact extrapolation technique they left the 12ug/m^3 recommendation intact as a one in one million risk.

(3) Does benzene cause hematological neoplasms other than ANLL?

One of the limitations in studying the effects of benzene in laboratory animals has been the difficulty in developing an animal model of benzene-induced ANLL. However, studies in laboratory animals have clearly demonstrated that benzene causes hematological neoplasms other than ANLL as well as non-hematological neoplasms (Maltoni, 1983; NTP, 1984; Snyder et al, 1982). This has naturally raised the question of whether benzene can cause cancers other than ANLL in humans (Savitz and Andrews, 1997; Goldstein, 1990; Goldstein and Witz, 2000). I believe that the answer is most likely yes, but still scientifically unproven, for a variety of hematological tumors including non-Hodgkins lymphoma (NHL), multiple myeloma and acute lymphatic leukemia (ALL).

In each of these three tumors derived from lymphocytic cells there is some epidemiological support, although controversial, as well as a strong element of biological plausibility. Lymphocytic cells are particularly at risk to benzene toxicity, the lymphocyte count decreasing even more rapidly than does the granulocytic count. Further, chromosomal effects are readily observed in the lymphocytes of humans and animals with significant benzene exposure. And longer term exposures to benzene causes lymphomas in laboratory animals.

This is not the place to enter into the details of the controversy concerning the epidemiology. Briefly, Wong and Raabe and their colleagues have recently published two large meta analyses in which they report no increased incidence of either multiple myeloma or of NHL (Bergsagel et al, 1999; Wong and Raabe, 2000). Both are seriously flawed by the fact that the populations under study do not appear to have had a statistically significant increased incidence of ANLL, reflecting the fact that benzene exposure for most workers in the petroleum industry is relatively well controlled and that many workers in these cohorts are at little risk of benzene exposure (Goldstein and Shalat, 2000; Bergsagel et al, 2000). It is unreasonable to ask the question of whether benzene can cause NHL or multiple myeloma in a cohort in which there is not a clear signal of benzene induced ANLL. The fallacy is similar to asking whether cigarette smoking can cause NHL in a large cohort whose level of cigarette smoking is too low to cause a statistically significant increase in risk of lung cancer.

An additional problem with their approach is exemplified by the findings of Rushton and Alderson (1981). Their nested case control study of leukemia in British petroleum workers showed a positive association between workplace benzene exposure and leukemia despite an overall SMR for leukemia

of 0.95. Wong et al (1999) have published a nested case control study of leukemia, acute myelogenous leukemia, multiple myeloma and kidney cancer in a cohort of petroleum workers exposed to gasoline. As they did not find an increased risk of acute myelogenous leukemia or of all leukemias in relation to benzene exposure, their negative findings for kidney cancer and multiple myeloma are simply not relevant to the issue of whether benzene can cause these latter two cancers. ALL is primarily a disease of children. While children are exposed to community sources of benzene, their absence from the workforce precludes usual epidemiological approaches to the question of whether benzene can cause ALL. About the most that we can reasonably conclude at present is that it is highly unlikely that the potency of benzene in producing ANNL is exceeded by its potency in producing any other human cancer.

A major issue in all of the extrapolation approaches from the pliofilm cohort is the lack of sensitivity due to the relatively small numbers involved. These small numbers also impact on the usual epidemiological approaches to determine if causality exists. For example, the four cases of multiple myeloma observed in the initial pliofilm study (Rinsky, 1981) were not preferentially observed in the more highly exposed work categories. But with only four myeloma cases, with one expected, it is hard to put much credence in the lack of a dose-related distribution. Similarly, Wong and Raabe (2000) have dismissed the finding of Consonni et al (1999) in which five cases of NHL were observed (2.12 expected) because of a lack of a statistically significant upward trend.

To summarize the above, in my judgment it is very likely but scientifically unproven that benzene causes hematological neoplasms other than NHL. A reasonable assumption is that this might lead to a doubling of the overall cancer risk.

Risk assessment for non-cancer hematological effects

There is no question that benzene causes hematological effects other than cancer. High level workplace exposures to benzene usually lead to more deaths from aplastic anemia than from ANLL. Further, in large well studied cohorts in which there have been cases of aplastic anemia, there are usually many more cases of pancytopenia with all degree of gradation from very mild to highly significant. Clinical manifestations other than the laboratory findings include symptoms due to anemia, an increased risk of infection due to a low white blood count, and an increased risk of hemorrhage due to a low platelet count. In attempting to estimate the health costs, it should be noted that there is a wide gap between the lower end of the statistically normal range of blood count values and the much lower blood counts that are required for symptoms or for overt clinically recognizable disease consequences that could be readily used for economic analysis. An individual with a mild to moderate benzene-induced pancytopenia will likely be clinically unrecognizable unless he or she happens to have routine blood counts.

For the purposes of the present exercise I will briefly attempt to distinguish among three levels of benzene induced non-cancerous effect: 1) the exposure level that will produce a clinically recognizable endpoint such as symptomatic anemia, infection or hemorrhage; (2) the exposure level that will produce a lowering of blood count(s) below normal levels; and (3) the level of benzene that might have any detectable hematological effects.

(1) Symptomatic effects: Such effects undoubtedly require benzene exposures well above the 10 ppm (32 mg/m^3) TWA workplace standard in effect for a few decades in the United States and elsewhere. Literally millions of workers were subjected to routine blood counts on a quarterly to annual basis. The reason the OSHA standard was decreased to its present 1 ppm level was solely because of cancer concerns, not because the higher standard was leading to non-cancer hematological disease. Recent data from China describe clinical aplastic anemia in factories with exposure levels said to range from 93-1156 mg/m^3 (Yin et al, 1987b). A not unreasonable assumption is that clinically overt symptoms will not occur as a result of long term benzene exposure to levels at the workplace below 100 mg/m^3, or perhaps much higher.

(2) Blood count(s) below the normal range: There are a number of studies evaluating blood counts in workers exposed to reasonably well defined levels of benzene, although in each study there is some grounds for uncertainty as to whether the measured benzene levels are pertinent to the specific individuals with low blood counts. In most cases the availability of blood count data is related to surveillance of benzene-exposed workers. Complicating interpretation of these blood count data is the fact that there are many reasons for variations in blood counts below the statistically normal value, some related to normal biological variation, some to laboratory variation, and some to the many other causes of low blood counts for reasons as diverse as viral infections and alcoholism.

Perhaps the most extensive study of workers exposed to benzene is that of Yin et al (1987b) who found 2,676 cases of benzene poisoning, defined as a white blood count less than 4,000/mm^3, in a review of over 500,000 benzene-exposed workers in China. The geometric mean concentration in 50,255 workplaces was 18.1 mg/m^3, and 64.6% of the workplaces had less than 40 mg/m^3. From their review of the data the authors conclude that cases of benzene poisoning may occur even in factories with less than 40 mg/m^3 benzene.

Gross chromosomal abnormalities in association with overt benzene hematotoxicity were originally reported by Forni et al (1971; see also Forni, 1996). More recent findings of chromosomal abnormalities using fluorescent in situ hybridization technique have been reported in Chinese workers with benzene exposure above 31 ppm (99 mg/m^3; Zhang et al, 1996).

(3) Any detectable hematological effects: Sensitive indicators of bone marrow toxicity have been explored in animal studies aimed primarily at determining the mechanism of benzene hematotoxicity. Mice are more sensitive to rats. Green et al (1981) looked at specific progenitor bone marrow cells in mice and reported effects at inhalation exposure levels of 9.9 ppm (32 mg/m^3), but not 1.1 ppm (3.5 mg/m^3) benzene, 6 hours per day for 5 days. Farris et al (1997) evaluated similar endpoints in mice. They found effects at 100 and 200 ppm (320 and 640 mg/m^3), 6 hrs a day for up to 8 weeks, but not at 1,5, and 10 ppm (3.2, 16 and 32 mg/m^3). Cytogenetic endpoints also have been evaluated in laboratory animals.

There are a number of worker studies that have reported statistically significant changes in hematological endpoints at relatively low benzene exposure levels. Ward et al (1996) reported on the blood counts of the pliofilm workers and suggested that there may be no threshold for the hematologic effects of benzene which may occur even at levels less than 5 ppm (16 mg/m^3). On the

other hand, Tsai et al (1998), based upon the lack of evidence of effects in hematological monitoring results from 2475 employees at a petrochemical complex, questioned the need for this form of surveillance. Khuder et al (1999) reported on a group of 105 petroleum workers exposed to 0.14 -2.08 ppm (0.45 - 6.6 mg/m^3) benzene who over time had small but statistically significant falls in certain blood counts. However, there were problems with this study, including a decrease in the red cell mean corpuscular volume, a finding contrary to what is observed in benzene toxicity (Goldstein and Cody, 2000). Nilsson et al (1996) reported findings suggestive of genotoxic effects in men occupationally exposed to relatively low levels of benzene in the range of 0.1 ppm (0.3 mg/m^3), but there were other exposures as well. Multiple exposures also is a confounding factor in the report of Carere et al (1995) of cytogenetic changes in Rome gasoline station attendants. There were also inconsistencies in relation to benzene exposure levels.

In summary, there is no convincing evidence of any non-neoplastic hematological effects at benzene levels in the range of current community air pollution levels.

Susceptibility issues

There is ample indirect evidence, as well as some direct evidence, of differences in susceptibility to the hematological effects of benzene among individuals. Women are believed to be more susceptible than men due to an average higher body fat leading to more benzene storage. There are genetic polymorphisms governing the activity of many of the known steps in the benzene metabolic pathway. One of the more intriguing recent findings in the field of genetic polymorphisms is the observation by Rothman et al (1997) that the risk of decreased blood counts among Chinese workers exposed to benzene increased seven-fold if the workers had two different phenotypic variations that led on the one hand to increase the rate at which initial benzene metabolism produced hydroxylated intermediates and on the other hand slowed the rate of detoxification of these metabolites. Workers with only one of these variants had approximately a doubling of risk. There is also the suggestion in the older literature that individuals with thalassemia were at increased risk of benzene toxicity, an observation that is perhaps generalizable to any groups with increased bone marrow precursors due to inherited anemias, e.g., sickle cell disease. Much of this work needs to be followed up before it can be used in economic analyses.

References

Benzene '95. (1995) An International Conference on the Toxicity, Carcinogenesis and Epidemiology of Benzene, Piscataway, NJ. *Environmental Health Perspectives*. 104: (Supplement 6).

Bergsagel, D.E., Wong, O., Bergsagel, P.L., Alexanian, R., Anderson, K.C., Kyle, R.A., and Raabe, G.K. (1999) Benzene and multiple myeloma: Appraisal of the scientific evidence. *Blood*, 95(4): 1174-1182.

Bergsagel, D.E., Wong, O., Bergsagel, P.L., Alexanian, R., Anderson, K.C., Kyle, R.A., and Raabe, G.K. (2000). Benzene and multiple myeloma: Scientific evidence. *Blood*, 95(4): Letter.

Caprino, L., and Togna, G.I. (1998). Potential health effects of gasoline and its constituents: A review of current literature (1990-1997) on toxicological data. *Environmental Health Perspectives*. 106(3): 115-125.

Carere, A., Antoccia, A., Crebelli, R., Degrassi, F., Fiore, M., Iavarone, I., Isacchi, G., Lagorio, S., Leopardi, P., and Marcon, F. (1995) Genetic effects of petroleum fuels: cytogenetic monitoring of gasoline station attendants. *Mutation Research*. 332(1-2): 17-26.

Chen, H., and Eastmond, D.A. (1995a). Accelerated paper: Topoisomerase inhibition by phenolic metabolites: A potential mechanism for benzene's clastogenic effects. *Carcinogenesis*. 16(10): 2301-2307.

Chen, H., and Eastmond, D.A. (1995b). Synergistic increase in chromosomal breakage within the euchromatin induced by an interaction of the benzene metabolites phenol and hydroquinone in mice. *Carcinogenesis*. 16(8): 1963-1969.

Clayton, C.A., Pellizzari, E.D., Whitmore, R.W., Perritt, R.L., and Quackenboss, J.J. (1999) National Human Exposure Assessment Survey (NHEXAS): distributions and associations of lead, arsenic and volatile organic compounds in EPA Region 5. *Journal of Exposure Analysis and Environmental Epidemiology*. 9: 381-392.

Cody, R.P., Strawderman, W.W., and Kipen, H.M. (1993) Hematologic effect of benzene: Job-specific trends during the first year of employment among a cohort of benzene-exposed rubber workers. *J. Occup. Med*. 35(8): 776-782.

Consonni, D., Pesatori, A.C., Tironi, A., Bernucci, I., Zocchetti, C., and Bertazzi, P.A. (1999). Mortality study in an Italian oil refinery: Extension of the follow-up. *American Journal of Industrial Medicine.* 35: 287-294.

Cox, L.A. Jr. (1996) Reassessing benzene risks using internal doses and Monte-Carlo uncertainty analysis. *Environ. Health Perspect.* 104(Suppl 6): 1413-29.

Crump, K.S. (1994). Risk of benzene-induced leukemia: a sensitivity analysis of the pliofilm cohort with additional follow-up and new exposure estimates. *J.Toxicol.Environ.Health.* 42(2): 219-242.

Crump, K., and Allen, B. (1984). Quantitative estimates of risk of leukemia from occupational exposure to benzene. OSHA Docket H-059b, Exhibit 152 (Appendix B). Washington, D.C.

Dosemeci, M., Li, G., Hayes, R.B., Yin, S., Linet, M., Chow, W., Wang, Y., Jiang, Z., Dai, T., Zhang, W., Chao, X., Ye, P., Kou, Q., Fan, Y., Zhang, X., Lin, X., Meng, J., Zho, J., Wacholder, S., Kneller, R., and Blot, W. (1994). Cohort study among workers exposed to benzene in China: II. Exposure. *American Journal of Industrial Medicine*. 26(3): 401-411.

Eastmond, D.A., Smith, M.T., and Irons, R.D. (1987) An interaction of benzene metabolites reproduces the myelotoxicity observed with benzene exposure. *Toxicol. Appl. Pharmacol*. 91: 85-95.
Environmental Protection Agency. (1997). Carcinogenic effects of benzene: an update. EPA/600/P-97/001A 1-43.

Environmental Protection Agency. 40 CFR Part 61 National emission standards for hazardous air pollutants; benzene

emissions from maleic anhydride plants, ethylbenzene/styrene plants, and benzene storage vessels; proposed withdrawal of proposed standards, [AD-FRL-2523-6], 49 FR 8386, March 6, 1984.

Farris, G., Robinson, S.N., Gaido, K.W., Wong, B.A., Wong, V., Hahn, W.P., and Shah, R.S. (1997). Benzene-induced hematotoxicity and bone marrow compensation in B6C3F1 mice. *Fundamental and Applied Toxicology*. 36: 119-129.

Forni, A. (1996). Benzene-induced chromosome aberrations: A follow-up study. *Environmental Health Perspectives.* 104(Suppl 6): 1309-1312.

Forni, A.M., Cappellini, A., Pacifico, E., and Vigliani, E.C. (1971) Chromosome changes and their evolution in subjects with past exposure to benzene. *Arch. Environ. Health*. 23: 385-391.

Goldstein, B.D. (1977). Hematotoxicity in humans. *J. Toxicol. Environm. Health* (suppl.) 2:69-105.

Goldstein, B.D. (1985) Risk assessment and risk management of benzene by the Environmental Protection Agency. *Banbury Report 19: Risk Quantitation and Regulatory Policy.*

Goldstein, B.D. (1990) Is exposure to benzene a cause of human multiple myeloma? *Annals of the New York Academy of Science*, 609:225-334.

Goldstein, B.D.and Cody, R. (Letter) Assessment of complete blood count variations among workers exposed to low levels of benzene. *J Occup Environ Med*. 42:2-113-114, February, 2000.

Goldstein, B.D.and Shalat, S.L. (2000) The causal relation between benzene exposure and multiple myeloma. *Blood*, 95(4): 1512-4. Letter

Goldstein, B.D. and Witz, G. Chapter 4: Benzene. In: Environmental Toxicants: Human Exposures and Their Health Effects. Second Edition. Ed. M. Lippman, John Wiley and Sons, Inc., pp.121-150, 2000.

Green, J.P., Snyder, C.A., Lo Bue, J., Goldstein, B.D. and Albert, R.E. (1981). Acute and chronic dose-response--Effects of inhaled benzene on multi potential hematopoietic stem (CFU/S) and granulocyte/macrophage progenitor (GM/CFU/C) cells in CD/1 mice. *Tox. Appl. Pharm.* 58:492-503.

Guy, R.L., Hu, P., Witz, G., Goldstein, B.D. and Snyder, R. (1991) Depression of iron uptake into erythrocytes in mice by treatment with the combined benzene metabolites p-benzoquinone, muconaldehyde and hydroquinone. *J. Appl. Toxicol.* 11: 443-446.

Hayes, R.B., Yin, S.-N., Dosemeci, M., Li, G.-L., Wacholder, S., Travis, L.B., Li, C.-Y., Rothman, N., Hoover, R.N., and Linet, M. (1997). Benzene and the dose-related incidence of hematologic neoplasms in China. *Journal of the National Cancer Institute*. 89: 1065-1071.

Health Council of the Netherlands: Committee on the Evaluation of the Carcinogenicity of Chemical Substances. (1997). Benzene. 1-44.

Infante, P.F., Rinsky, R.A., Wagoner, J.K., Young, R.J. (1977). Leukemia in benzene workers. *Lancet*. I: 76-78.

Irons, R.D. (2000) Molecular models of benzene leukemogenesis. In: Special issue on benzene. *Envir. Health Perspect.* (Supplement to be published). Krewski, D., and Snyder, R. (editors).

Johnson, T.R., Roy, A.P., and Capel, J.E. (1991) Estimation of incremental benzene exposure associated with seven bulk gasoline storage facilities in North Carolina. Health and Environmental Sciences Department, *American Petroleum Institute*. Report, 1-36.

Kipen, H.M., Cody, R.P., Crump, K.S., Allen, B. and Goldstein, B.D. (1988) Hematologic effects of benzene: A thirty-five year longitudinal study of rubber workers. *Toxicology & Industrial Health*, 4:411-430.

Khuder, S.A., Youngdale, M.C., Bisesi, M.S, Schaub, E.A. (1999) Assessment of complete blood count variations among workers exposed to low levels of benzene. *J. Occup. Environ. Med.* 41:821-826.

Krewski, D., and Snyder, R. (editors). (2000) Special issue on benzene. *Envir. Health Perspect.* (Supplement).

Lagrone, F.S. (1989). Assessing community exposure to selected volatile indicator compounds. *Journal of the American College of Toxicology*. 8(5): 897-904.

Laskin, D.L. (2000) Nitric oxide as a mediator of benzene-induced hematosuppression and toxicity. In: Special issue on benzene. *Envir. Health Perspect.* (Supplement to be published). Krewski, D., and Snyder, R. (editors).

Levay, G. (1992) Potentiation of DNA adduct formation in HL-60 cells by combinations of benzene metabolites. *Proceedings of the National Academy of Science of the United States of America*. 89: 7105-7109.

Maltoni, C., Conti, B. , and Cotti, G. (1983). Benzene: A multipotential carcinogen. Results of long-term bioassays performed at the Bologna Institute of Oncology. *American Journal of Industrial Medicine*, 4: 589-630.

Mani, C., Freeman, S., Nelson, D.O., Vogel, J.S., and Turteltaub, K.W. (1999) Species and strain comparisons in the macromolecular binding of extremely low doses of [14C] benzene in rodents, using accelerator mass spectrometry. *Toxicol. Appl. Pharmacol*. 159(2):83-90.

Nilsson, R.I., Nordlinder, R.G., Tagesson, C., Walles, S., and Jarvholm, B.G. (1996). Genotoxic effects in workers exposed to low levels of benzene from gasoline. *American Journal of Industrial Medicine*, 30: 317-324.

National Toxicology Program. (1984). NTP Technical Report on the Toxicology and Carcinogenesis Studies of Benzene (CAS #71-43-2) NF344/N Rats and B6C3F$_1$, Mice (Gavage Studies) NTP TF 289. U.S. Department of Health, Education and Welfare.

Paustenbach, D.J., Bass, R.D., and Price, P. (1993). *Environmental Health Perspectives Supplements*, 101(Suppl. 6): 177-200.

Paustenbach, D.J., Price, P.S., Ollison, W., Blank, C., Jernigan, J.D., Bass, R.D., and Peterson, H.D. (1992) Re-evaluation of benzene exposure for the plioform (rubberworker) cohort (1936-1976). *Journal of Toxicology and Environmental Health*, 36: 177-231.

Paxton, M.B., Chinchilli, V.M., Brett, S.M., and Rodricks, J.V. (1994a). Leukemia risk associated with benzene exposure in the pliofilm cohort: I. Mortality update and exposure distribution. *Risk Analysis*, 14(2): 147-154.

Paxton, M.B., Chinchilli, V.M., Brett, S.M., and Rodricks, J.V. (1994b). Leukemia risk associated with benzene exposure in the pliofilm cohort: II. Risk estimates. *Risk Analysis*, 14(2): 155-161.

Rinsky, R., Smith, A.B., Hornung, R., Filloon, T.G., Young, R.J., Okun, A.H., and Landrigan, P.J. (1987). Benzene and leukemia: An epidemiologic risk assessment. *New England Journal of Medicine.* 316: 1044-1050.

Rinsky, R.A., Young, R.J., Smith, A.B. (1981) Leukemia in benzene workers. *Am J. Industrl. Med*. 2:217-245.

Rothman, N., Hayes, R.B., Haas, R. et al. (1995). Benzene induces gene-duplicating but not gene-inactivating mutations at the Glycophorin A locus in exposed humans. *Proc. Natl. Acad. Sci. USA*, 92:4069-4073.

Rothman, N., Li, G., Dosemeci, M., Bechtold, W.E., Marti, G.E., Wang, Y., Linet, M., Xi, L., Lu, W., Smith, M.T.,

Titenko-Holland, N., Zhang, L., Blot, W., Yin, S., and Hayes, R.B. (1996). Hematotoxicity among Chinese workers heavily exposed to benzene. *American Journal of Industrial Medicine*, 29:236-246.

Rothman, N., Smith, M.T., Hayes, R.B., Traver, R.D., Hoener, B., Camleman, S., Dosemeci, M., Linet, M., Zhang, L., Xi, L., Wachholder, S., Lu, W., Meye, K.B., Titenko-Hollanad, N., Stewart, J.T., Yin, S., and Ross, D. (1997) Benzene Poisoning, A Risk Factor for Hematological Malignancy, is Associated with the NQ01 609C–>T Mutation and Rapid Fractional Excretion of Chloroxazone. *Cancer Res*. 15:2839-2842.

Savitz, D., and Andrews, K.W. (1997). Review of epidemiologic evidence on benzene and lymphatic and hematopoietic cancers. *American Journal of Industrial Medicine*, 31(3): 287-295.

Schnatter, A.R., Nicolich, M.J., and Bird, M.G. (1996). Determination of leukemogenic benzene exposure concentrations: Refined analyses for the pliofilm cohort. *Risk Analysis*, 16(6): 833-840.

Smith, M.T., and Fanning, E.W. (1997). Meeting Report: Report on the Workshop Entitled: "Modeling Chemically Induced Leukemia - Implications For Benzene Risk Assessment. *Leukemia Research*, 21(5): 361-374.

Smith, M.T., and Rothman, N. (2000) Biomarkers in the molecular epidemiology of benzene-exposed workers. In: Special issue on benzene. *Envir. Health Perspect.* (Supplement to be published). Krewski, D., and Snyder, R. (editors).

Snyder, C.A., Goldstein, B.D., Sellakumar, A., Bromberg, I., Laskin, S. and Albert, R.E. (1982) Toxicity of chronic benzene inhalation: CD-1 mice exposed to 300 ppm. *Bull. Environm. Contam. Toxicol*. 29:385-391.

Snyder, R., and Hedli, C.C. (1996) An overview of benzene metabolism. *Environ. Health Perspect. Suppl*., 104:1165-1171.

Snyder, R., Witz, G. and Goldstein, B.D. (1993) The toxicology of benzene. *Environmental Health Perspectives*, 100:293-306.

Tsai.S.P., Bennett, J.M., Salesman, C.N., Ryan, T.E., Gilstrap, E.L., and Ross, C.E. (1998). Medical surveillance for hematological disorders among active and retired refinery workers. *Journal of Environmental Medicine*. 40(5): 475-480.

Utterback, D.F. and Rinsky, R.A. (1995). Benzene exposure assessment in rubber hydrochloride workers: A critical evaluation of previous estimates. *American Journal of Industrial Medicine*, 27:661-676.

Wallace, L.A. (1986). The California Team Study: Breath concentrations and personal exposures to 26 volatile compounds in air and drinking water of 188 residents of Los Angeles, Antioch, and Pittsburg, CA. *Atmospheric Environment*, 22(10): 2141-2163.

Wallace, L.A. (1991). Comparison of risks from outdoor and indoor exposure to toxic chemicals. *Environmental Health Perspectives*, 95: 7-13.

Ward, E., Hornung, R., Morris, J., Rinsky, R., Wild, D., Halperin, W., and Guthrie, W. (1996). Risk of low red or white blood cell count related to estimated benzene exposure in a rubberworker cohort (1940 - 1975). *American Journal of Industrial Medicine*, 29: 247-257.

Wong, O. (1998). Re: Benzene and the dose-related incidence of hematologic neoplasms in China. *J. Nat. Cancer Inst*. 90: 469-471.

Wong, O. (1999). A critique of the exposure assessment in the epidemiologic study of benzene-exposed workers in China conducted by the Chinese Academy of Preventive Medicine and the US National Cancer Institute. *Regul. Toxicol. Pharmacol*. 30: 259-267.

Wong, O., and Raabe, G.K. (1989). Critical review of cancer epidemiology in petroleum industry employees, with a quantitative meta-analysis by cancer site. *American Journal of Industrial Medicine*. 15: 283-310.

Wong, O., and Raabe, G.K. (1995) Cell-type-specific leukemia analyses in a combined cohort of more than 208,000 petroleum workers in the United States and the United Kingdom, 1937-1989. *Regul. Toxicol. Pharmacol*. 21(2): 380-4.

Wong, O., and Raabe, G.K. (2000). Non-Hodgkin's lymphoma and exposure to benzene in a multinational cohort of more than 308,000 petroleum workers, 1937 to 1996. *JOEM*. 42(5):554-567.

Wong, O., Trent, L., and Harris, F. (1999) Nested case-control study of leukemia, multiple myeloma, and kidney cancer in a cohort of petroleum workers exposed to gasoline. *Occup. Environ. Med.* 56:217-221.

Woodruff, T.J., Axelrad, D.A., Caldwell, J., Morello-Frosch, R., and Rosenbaum, A. (1998) public health implications of 1990 air toxics concentrations across the United States. *Environmental Health Perspectives*. 106(5):245-251.

Xu, X., Wienchke, J.K. , Niu, T., Wang, M., Watanabe, H., Kelsey, K.T., and Christiani, D.C. (1998). Benzene exposure, Glutathione S-Transferase Theta homozygous deletion, and sister chromatid exchanges. *American Journal of Industrial Medicine*. 33: 157-163.

Yin, S.-N., Li, G.-L., Tain, F.-D., Fu, Z.-I., Jin, C., Chen, Y.-J., Luo, S.-J., Ye, P.-Z., Zhang, J.-Z., Wang, G.-C., Zhang, X.-C., Wu, H.-N., and Zhong, Q.-C. (1987a). Leukemia in benzene workers: a retrospective cohort study. *British Journal of Industrial Medicine*. 44: 124-128.

Yin, S.-N., Li, Q., Liu, Y., Tain, F., Du, C., and Jin, C. (1987b). Occupational exposure to benzene in China. *British Journal of Industrial Medicine*. 44:192-195.

Yin, S.-N., Hayes, R.B., Linet, M.S., Li, G.-L., Dosemeci, M., Travis, L.B., Zhang, Z.-N., Li, D.-G., Chow, W.-H., Wacholder, S., Blot, W.J., and The Benzene Study Group. (1996a). An expanded cohort study of cancer among benzene-exposed workers in China. *Environmental Health Perspectives*. 104(Supplement 6): 1339-1341.

Yin, S., Hayes, R.B., Linet, M.S., Li, G., Dosemeci, M., Travis, L.B., Li, C., Zhang, Z., Li, D.-G., Chow, W., Wacholder, S., Wang, Y., Jiang, Z., Dai, T., Zhang, X., Lin, X., Meng, J., Ding, C., Zho, J., and Blot, W. (1996b). A cohort study of cancer among benzene-exposed workers in China: overall results. *American Journal of Industrial Medicine*. 29: 227-235.

Zhang, L., Rothman, N. , Wang, Y., Hayes, R.B., Bechtold, W., Venkatesh, P., Yin, S., Dosemeci, M., Li, G., Lu, W., and Smith, M.T. (1996). Interphase cytogenetics of workers exposed to benzene. *Environmental Health Perspectives*. 104(Supplement 6): 1325-1329.

APPENDIX F-3

Challenges in Projecting Human Health Impacts from Exposures to Perchloroethlyene

White Paper by Dr. Lorenz Rhomberg,
Gradient Corporation, Cambridge, MA

Challenges in Projecting Human Health Impacts from Exposures to Perchloroethylene

Lorenz R. Rhomberg, Ph.D.

Principal Scientist
Gradient Corporation
Cambridge, MA

1 Introduction

When analysis of the toxic effects of chemicals is applied to the task of assessing benefits of regulations that limit exposure, the goal is to estimate the impact of changes in exposure regimes on changes in the burden of ill health in the exposed population. This differs from the aim of traditional regulatory risk assessment, which is to define exposure levels that can be deemed "safe," or at least that can be found to pose no more than "acceptable" risks. That is, the usual methods seek to define dose levels without pronounced impacts, not to estimate or to characterize the impacts that may occur.

Not surprisingly, traditional methods are ill suited to the estimation and description of toxic effects to be expected when chemical exposures approach and exceed levels that can be assuredly ruled safe. An often-mentioned issue is that traditional risk assessment methods are "conservative" in that they deal with uncertainties in the inferential process by making estimates or assumptions unlikely to underestimate risk, thereby tending to overestimate risk, at least on average. Such biases distort the assessment of benefits gained from avoided exposure.

Two further issues are perhaps as important, however, and may be more difficult to remedy. First, owing to their focus on defining doses without unacceptable effects, existing approaches often say little about what specific toxic phenomena are to be expected at exposures exceeding "safe" levels. Second, because they focus on individual risks to benchmark individuals with defined "high-end" exposures, existing approaches are not geared to estimating population risks among a large group of subjects with varying levels of exposure that may fluctuate in time or consist of occasional high-exposure episodes.

To undertake benefits assessment, some new approaches to analysis of toxic effects will have to be considered. This paper attempts to examine the challenges and to consider modifications to risk analysis methods that may help to address some of the questions. To ground the discussion in the context of data that are available for actual toxic agents, the example of perchloroethylene is used.

It is best to begin by defining goals, even if they represent ideals that we are unlikely to achieve in practice. In order for economic analysis to measure the benefits of regulations that restrict exposure to a potentially toxic agent, it must have estimates of the burden of ill health in the exposed population as it would be expected to exist when the regulation is applied as well as when it is not applied. The differences between these constitute the avoided health impacts, and the values placed on them (which, thankfully, it is not my task to address) largely constitute the benefits of the regulation. Clearly, at least one of these scenarios (with the regulation or without) is hypothetical, and so even in the ideal case we cannot rely solely on observation. Modeled projection of health impacts to be expected in a population under hypothetical exposure scenarios is a necessary part of the analysis.

1.1 Needs

It would seem that the ideal toxicological analysis would provide characterization of the following:

1. *What* specific responses are engendered by exposures to the toxic agent?
2. For responses that are graded, *how severe* is the response? How does severity progress over time?
3. *When* (in the course of an ongoing exposure) do responses arise? *How long* do ill effects endure?
4. *In whom* do responses arise?

These bear some discussion. First, to put value on a case of toxicity avoided, it helps to specify the effect in question, since different effects (and different severities) have different impacts on the quality of life. Exiting methods typically eschew making statements about the specific nature of toxic effects in humans that are extrapolated from animal studies. Animal carcinogenicity is assumed to indicate a human risk for some type of cancer, but this is not necessarily expected to manifest itself as the same type of cancer seen in the animals. Noncancer toxicity assessments define doses that appear to avoid all adverse responses seen among experimental animals, and the most sensitive of these is deemed the "critical effect," but it is not specified which effects are to be expected in humans in exposures that exceed "safe" levels. Ideally, then, methods for benefits assessment should aim at making more specific projections about the nature of the toxicity to be expected in sufficiently exposed humans. They should recognize that several toxic effects may be at issue, not solely the one that was used to set the acceptable dose in the regulation being examined.

Similarly, existing methods do not project when during the course of an ongoing exposure the adverse effects will become manifest. To place a value on a case of toxicity, however, one would want to know when in life it appears, how long the state of ill health endures, whether it changes in severity, whether the disease fully or partly regresses upon cessation of exposure, how much the length of life is shortened,

and how a change in exposure at some midlife point (resulting, say, from the imposition of a regulation) changes the likelihood of response. In short, one would like not just dose-response relationships, but descriptions of response as a function of dose-rate and time, including description of the consequences of non-constant dose rates. Exposure-dependence of the course of disease in any cases engendered is also of interest.

It should be clear that the concern is for population risks, not just individual risks (which are often the focus of traditional assessments). We seek to characterize all the effects as they are (or would be) realized in an actual human population of interest. The hypothetical 70-year fence-line resident, the pica child, the worker laboring 45 years at a degreasing tank, and other standardized individual scenarios of exposure that define benchmarks of individual risk in regulation-setting assessments are not at issue, not just because they are "high-end" exposures but because they represent but a few individuals among the many in the population whose collective benefits we wish to address.

Ideally, we would want to describe not only the full frequency distribution of exposure levels, but also when and by whom the various exposures are experienced, since exposures at different ages or in different patterns over lifetime will differently affect the likelihood of responses (and we may wish to place different value on responses occurring at different times or in people with different prior states of health). Multiple sources and pathways of exposure exist, and people change their geographic locations and local exposures on timescales ranging from minutes to years. When a regulation is phased in, or when an agent persists in the environment even after controls are imposed, the exposures will change year by year, and this time pattern may be important to characterize to gauge accrual of benefits from the exposure restriction.

These facts make for major challenges to exposure assessment. (Exposure methods are not my focus, but the issues should not be overlooked.) Creating a complete inventory of the individual histories of exposure in an entire diverse population may seem a daunting task, but considerable progress has been made in approaching such a description using simulation modeling. In this approach, the events and settings that lead to exposure in a population are described as random variables, and a large set of simulated life histories can be assembled (ILSI 1998) that describes the diversity of experiences in a whole population.

1.2 Uncertainty

Existing methods in risk assessment for projecting human risks from experimental observations of toxicity in animals are highly uncertain. Even use of epidemiological studies entails uncertainty in characterization of exposures, in description of responses, and in generalization from the study population to the more general population of interest. New methods that attempt to make more detailed pronouncements regarding the nature of endpoints and the timecourse of their manifestation while acknowledging the complexity of the distribution of human exposures are bound to be still more uncertain.

Any demand that an analysis of benefits cannot be undertaken until impacts of exposure can be projected with confidence dooms the enterprise. It also misses the point. While we do our best to project outcomes with precision, uncertainty cannot be avoided, only characterized. It is the fact that outcomes are uncertain that makes them risks. The assessment of the costs and benefits of regulation can be regarded as a problem in decision under uncertainty—we have to decide how much to spend to control exposures in the face of uncertainty about how much benefit (in terms of reduced health impact) we will receive. The decision to incur regulatory costs is deemed a good one for society if the mathematical expectation of

the uncertain benefits exceeds that of costs. The expectation is not the single most likely value, but rather the average over possibilities, each weighted by its likelihood of being true.

Seen from this point of view, the characterization of uncertainty in the projection of health effects is central to the analysis. The characterization of risk consists of specifying an array of possible outcomes or courses of events, each element of which is associated with the likelihood of its occurrence and the consequences should it indeed occur (Kaplan and Garrick 1981). In the present case, the likelihoods constitute our relative confidence in the alternative projections of health impacts. We want to avoid the "upper bound" and "worst-case" nature of much of existing methodology, but at the same time we should not seek only single "best estimates" (such as maximum likelihood curve fits). Instead, we should seek to characterize the distribution of possibilities.

2 Perchloroethylene

Perchloroethylene ("Perc," CAS No.127-18-4) is a high production-volume chlorinated solvent used as a chemical intermediate, as a solvent and degreasing agent, and as the primary solvent in drycleaning operations. Perchloroethylene is moderately volatile; without containment and measures for vapor recovery, use and disposal can result in considerable release of vapor to the atmosphere. Spills and leaks during storage have resulted in cases of contaminated soil and groundwater. Because the resulting exposures to workers and the general public lead to concerns for potential human health effects, the use and disposal of perchloroethylene is subject to regulation aimed at limiting workplace concentrations and releases of the chemical to the environment. The mandated controls can be costly, and it is of interest to establish how much impact on the health of the human population is avoided through their application.

A full review of the exposures to perchloroethylene and a complete characterization of its toxicologic and epidemiologic database are beyond the scope of this paper. The following overview, drawn from IARC (1995), EPA (1985, 1991), ATSDR (1997), OEHHA (1999) and other sources, gives a perspective on the available information that is sufficient for the present discussion of methodological issues.

Exposure: Worldwide annual production of perchloroethylene (which has declined somewhat in recent years) is in the hundreds of thousands of tons. Some 55% is used for drycleaning, 23% as a chemical intermediate (mostly for CFC production, which is declining), and 13% for liquid and vapor degreasing, with other uses including fabric treatment and paint stripping. Sampled air concentrations vary considerably in degreasing facilities, but means are often on the order of 10-100 ppm (parts per million) with some individual air samples in the 1,000 ppm range. Occupational exposures in drycleaning facilities are on the order of 10-50 ppm (IARC 1995).

Ambient air levels are much less and are reported here in parts per billion (1000 ppb = 1 ppm); they vary somewhat with season and are generally higher in urban than in rural air. Levels of 0.2 to 2 ppb are usually found in outdoor urban air. Indoor levels are often higher, sometimes tenfold outdoor levels. Peak levels in apartments above drycleaning establishments have been measured at 1000 ppb and higher. Off-gassing from drycleaned clothes can lead to temporarily high levels in automobiles (about 2000 ppb but with reports up to 300,000) and in homes (about 400 ppb).

Pharmacokinetics and Metabolism: Perchloroethylene is readily absorbed after inhalation or ingestion. Much of this is exhaled unchanged, but on the order of 30-50% is metabolized at low exposure levels in both rodents and humans. Most of this metabolism is *via* an oxidative pathway, but at exposure levels higher than about 100 ppm (such as in rodent lifetime bioassays), the oxidative pathway becomes increasingly saturated. This leads to proportionally higher metabolism by a glutathione-conjugation

pathway (which is still small in absolute terms) and higher exhalation of unmetabolized compound. Glutathione conjugates can be further metabolized in kidney to reactive, apparently genotoxic compounds, but oxidative metabolites (and perc itself) do not appear to be genotoxic. Several pharmacokinetic models of perchloroethylene metabolism exist; they agree in broad outline but differ in detail, especially regarding rodent-human differences in the extent of the conjugative pathway.

2.1 Observations of Toxicity in Humans

Neurological effects have been seen in populations occupationally exposed to bouts of high perchloroethylene concentration in air, and the nervous system seems to constitute the most susceptible target in humans. Overt symptoms such as headache, nausea, and ataxia are not seen in experiments at doses below 100 ppm, and these are fully reversible. Subtler pre-clinical neurophysiological and neurobehavioral effects such as changes in electroencephalograms, visual-evoked potentials, color vision discrimination, and tests of coordination or reaction time show detectable influence of exposure at levels between 15 and 100 ppm, although these, too, are reversible upon cessation of exposure. No clear evidence suggests permanent neurological effects from chronic occupational exposure, but some studies report detection of significant differences in memory and reaction time.

Case studies exist of workers exposed to very high levels in industrial accidents (*e.g.*, a worker found unconscious in a pool of solvent) in which serious liver or kidney damage occurred, but in such cases there is apparent full recovery within weeks. As with neurological effects, subtler pre-clinical changes that are considered markers of potential toxicity are seen in some studies of workers with exposures in the 20-30 ppm range. Many of these are elevations in serum concentrations of certain liver-cell enzymes (SGOT, SGPT, GGT) that are taken to signal some loss of integrity or increased permeability of liver cells, and hence possible beginnings of hepatotoxicity. It is typical for these quantitative measures to be within normal range in all subjects yet the means for exposed and unexposed groups are statistically different.

Some studies suggest slightly increased rates of spontaneous abortion or menstrual complaints in women with occupational exposures, and some studies suggest longer times to conception in couples with one or the other parent exposed. No associations with stillbirth, low birthweight, or malformations have been seen.

Several occupational epidemiological studies of carcinogenic effects have been conducted of drycleaning workers and those exposed in settings where degreasing activities lead to elevated air concentrations. Various inconsistent small elevations of one or another type of tumor have been reported (lymphopoietic, female genital, bladder, kidney, breast), but the only one showing any consistency is esophageal cancer. This effect (SMR 2.1 and 2.6) was seen in two drycleaning employee cohorts (but only in black men in one of them). A case-control study of esophageal cancer showed a non-significant association with employment in drycleaning. Esophageal cancer is subject to influence by smoking and alcohol use. Moreover, perchloroethylene is not the only chemical exposure for many of the workers in these studies.

2.2 Observations of Toxicity in Experimental Animals

Many of the noncancer effects seen in humans are seen in animals as well, but often at higher doses and for more overt and frankly toxic versions of the effect (since subtle effects are difficult to detect). Thus, animals acutely exposed to over 1000 ppm showed ataxia and anesthesia as well as altered psychomotor functions. Effects on brain weight were seen in rats at 600 ppm for 4 or 12 weeks. High exposures also produce liver and kidney toxicity, and the biochemical markers such as serum enzymes also appear at

exposures on the order of 25 ppm. Some effect on litter size and survival during lactation were seen at 1000 ppm.

In lifetime carcinogenicity bioassays, perchloroethylene by gavage (NCI 1977) and by inhalation (NTP 1986) increased hepatocellular carcinomas in male and female mice. The NCI study has been questioned because the perchloroethylene used was stabilized with epichlorohydrin, itself an animal carcinogen. Inhalation in rats led to an increase in mononuclear cell leukemia in both sexes, although response was no higher at the high than at the low dose. In addition, treated male rats had a few renal tubular cell adenocarcinomas that, although not statistically elevated compared to controls, were considered toxicologically significant owing to their historical rarity.

None of these animal tumor responses is without some controversy regarding its applicability as indicator of potential human risk. Mice of the strain tested are particularly prone to such tumors, and they appear at high levels even in controls. A major metabolite of perchloroethylene, trichloroacetic acid, induces proliferation of peroxisomes in mouse liver cells at high doses, and the damage or oxidative stress these cause may be involved in the induction of tumors, although other evidence questions the role of peroxisomes in hepatocarcinogenesis and the correlation of their induction with liver tumor induction has counterexamples. Humans have very little peroxisomes induction, even at high exposures, and the background rate of liver cancer is much lower than seen in mice. Meanwhile, trichloroacetic acid administered to mice in drinking water or experienced as a metabolite of trichloroethylene (which is similar in toxicology and metabolism to perchloroethylene) causes similar liver tumors at doses below those inducing peroxisomes and without inducing evident cell proliferation.

Similarly, the rat strain studied is prone to mononuclear cell leukemias, a tumor type with no clear analogue in humans (it is splenic, and human leukemias originate in marrow). The rat controls have high responses, although the rate is observed to vary among studies. Male rats can develop kidney tumors from some chemicals that inhibit degradation of a male rat-specific protein (α_{2u}-microglobulin) that accumulates in renal tubule cells, causing toxicity. This syndrome is unique to male rats and is considered irrelevant to human risk (since humans lack the mechanism altogether). Perchloroethylene metabolites appear to cause this phenomenon in male rats, but only at doses higher than those in the NTP bioassay, suggesting that a different mechanism is responsible. On the other hand, bioassay-level exposures to perchloroethylene do induce kidney toxicity, probably as a result of the kidney's ability to further metabolize products of the conjugative pathway into reactive compounds (which also may be genotoxic). But this phenomenon, including the kidney toxicity, is seen in mice as well, and mice do not have elevations in kidney tumor risk. There is evidence that the conjugative pathway and the activation of metabolites in kidney happen in humans, but the quantitative extent is unclear.

3 Projecting Cancer Risks

If we want to assess the benefits of limiting perchloroethylene exposure in terms of avoiding cancer risks, the first question to face is whether perc is a human carcinogen at all. One possible stance is to conclude that evidence is insufficient to treat this compound as a human carcinogen, and hence there is no cancer risk among exposed people (and thus no benefit from restricted exposure). Even if we feel that this is the single best-supported conclusion, however, there is some probability that we are wrong, and if we are, the cancer risk that may exist is overlooked. By the same token, it would be a mistake to put all our credence in an analysis that assumes that perc is a human carcinogen, ignoring the substantial probability that any risks so calculated are illusory.

Current methods force just such an either-or decision, with the decision process couched in the weight-of-evidence determination in hazard identification. In the case of perchloroethylene, the weight-of-evidence regarding human carcinogenicity is particularly muddled. IARC (1995) has called perc a 2B "probable human carcinogen" based on what it judges to be "limited" epidemiological evidence and "sufficient" animal evidence. EPA has withdrawn a former B2 classification, and the SAB has declared perc to be on the borderline between B2 and C. For the purposes of benefits assessment, our purpose should not be to resolve the hazard question, but to figure how best to hedge our estimates of cancer risks to account for the ambiguity. At present, there is no rigorous analytical scheme for doing this, so we need to rely on some kind of expert judgment. For sake of argument, I propose to put 10% weight on the possibility that perchloroethylene does indeed pose a human cancer risk (at some levels of exposure relevant to the assessment), and 90% on the possibility that it does not. My judgment attempts to account for the inconsistency of results among epidemiological studies, the likelihood that exposures to other agents or confounding by smoking or alcohol apply, the inconsistency among animal cancer results and lack of concordance with observations in humans, and the lack of biological hypotheses for why esophageal cancer in particular should be caused by perchloroethylene.

A variant of this approach would be to make separate judgments about each potential basis for a human cancer risk estimate, *i.e.*, a judgment about the esophageal cancer, about the bladder cancer, about the hematopoietic cancers, *etc.* seen among the human studies, as well as judgments about the liver cancer, leukemias, and kidney tumors in the animal studies. Each weight could then be multiplied by the study-specific estimate of risk (made contingent on its presumed relevance). This appropriately allows some (very small) probability that, say, both the mouse liver tumors and the human-study bladder tumors are indicating some actual human cancer risk from perc.

At this point it is probably wise to emphasize the distinction between using such a hedging approach for setting a regulation in the first place and for estimating the benefits of a regulation set by some other reasoning. In my view, some degree of conservatism and precaution in setting allowable exposures is legitimate. What we get for our money is not just the reduction of health impacts, but some degree of assurance that we have done enough to protect public health. Nonetheless, when the question is the estimation of what the regulation has accomplished, what is needed is our best attempt to make objective estimates of the relative likelihoods that various levels of benefit have been achieved. Such an analysis informs not only the expected benefits (the mean over possibilities) but also the assessment of how much assurance we have in fact achieved.

The next question is to ask what the cancer potency is in humans, contingent on our provisional consideration that there is one. The problem most often pointed to in this realm is that current methods for describing dose-response relationships define "upper bound" risks rather than central estimates. As noted previously, the solution is not to use the single best-fitting dose-response equation (the maximum likelihood estimate), since this fails to express the variety of more-or-less reasonable dose-response relations and does not in general reflect the expected value of the risk.

Instead, a useful approach is to conduct a bootstrap analysis of the dataset. In this simulation-based approach, a large number of alternative datasets are generated by resampling the original data (with replacement), and a best-fitting curve is generated for each iteration. This expresses the variation in low-dose potency to be expected as a result of the experimental error inherent in a limited number of observations, and the mean of the distribution gives an unbiased estimate of the expected value over the various possible values, with each possibility weighted by its likelihood of occurrence.

Although this kind of experimental error is what is allowed for in the upper-bound calculations of traditional methods, such error does not constitute the only, or even the primary, source of uncertainty in estimates of human low-dose cancer potency. There are many analytical choices made in the projection of animal-based risk estimates to humans. Notably, these include the choice of dose-response model to fit (which ideally should reflect understanding of underlying biological mechanisms of the agent's toxic action) and the means for determining the toxicologically equivalent exposures in the experimental animals and in humans. Such factors should be thought of as aspects of model uncertainty, since they reflect not alternative realizations of some underlying distribution, but rather our uncertainty as to what structure for the analytical approach gives the best projections.

Several studies have attempted to address this kind of uncertainty by an extension of the "hedging" approach described above. Each analytical choice is expressed as a stated set of alternatives, and the alternatives are then given weights to reflect the perceived relative plausibility of the approaches they embody. In a simulation approach, one can then iterate the analysis many times, each time choosing one of the alternatives for each factor with a likelihood proportional to the weights they have been assigned. The resulting distribution of outcomes gives a description of the array of possible overall analytical answers and their relative plausibility. McKone and Bogen (1992) applied this approach to perchloroethylene cancer risks from contaminated drinking water, although they gave equal weights to all the alternative datasets and analytical methods considered. Thompson and Evans (1997) built on this approach to consider cancer risks from perc use in drycleaning. A major advantage of such analysis is that it allows examination of the contribution to overall uncertainty from the various components, and it lends itself to value-of-information analysis that seeks to define how investment in research efforts to reduce key uncertainties can be expected to pay off in informing regulatory decisions. (They found that the expected value of perfect information about perchloroethylene's potency exceeds that about exposures.)

Evans *et al*. (1994) applied a more extensive version of this approach to the description of the carcinogenic potency of chloroform. They used a panel of experts to provide weights on the various analytical choices, and they allowed for the weights placed on alternatives for one factor to be contingent on choices for other factors. They found a wide but not unreasonable distribution of implied potencies. The then-existing EPA potency estimate fell at a high percentile of the estimates, as is appropriate for an upper bound, but the whole distribution provides perspective on the expected amount of benefit that limits on chloroform exposure could be thought to achieve.

This process of elaboration of possible alternatives could be drawn out indefinitely, so one has to devise an approach that captures the main sources of uncertainty and describes them adequately for the purposes at hand. In the case of perchloroethylene, we have several (poor) choices of datasets to analyze (and hence a large weight on the notion that none of them is applicable), several alternative pharmacokinetic models, each of which could be subject to a characterization of the uncertainty distribution of its estimates of values of several different dose measures (reflecting different perchloroethylene metabolites in different tissues, in mice, rats, and humans), with different dose-response approaches to be considered in view of judgments about mechanism of carcinogenic action. Clearly, the approach is not easy to implement, but simplified versions could be used to give a reasonable view of the uncertainty about projections of human cancer risk.

Once we have such projections, we need to deal with the fact that they are unspecific about the kind of cancer to be expected in humans as well as the time of appearance of any tumors that are in fact caused. At present, there is no very satisfactory method for specifying these, but it is worthwhile considering how important it really is to do so. If we assume that most cancers have roughly similar impact on length and

quality of life, and if we assume that induced cancers appear with the same distribution over ages as the general burden of background cancers, we will probably not be far off.

4 Projecting Noncancer Health Effects

Many of the issues just discussed regarding cancer risks apply to noncancer risks as well, but there are some additional questions to be considered.

First, "noncancer" toxicity is a catchall category, and a single chemical may cause several different kinds of noncancer effects. In the traditional assessment process, a critical effect is identified as the basis of setting an exposure below which no adverse responses are expected, but above such a level, various toxicities may be caused, and as doses increase, the number of endpoints that may become important may increase, as effects with higher and higher population thresholds come into play. For instance, moderately high doses of perchloroethylene may cause neurological effects, and still higher ones may cause these plus renal toxicity. We must therefore keep in mind that a series of parallel endpoint-specific assessments is necessary, and not just an assessment of the endpoint on which the RfD is based.

Second, noncancer endpoints vary considerably in their severity. This is always part of the debate about "adversity" that arises when one is defining the critical effect. A benefits assessment must consider the fact that avoidance of some effects that are not frankly adverse may nonetheless have some value (albeit less than might be ascribed to a more severe effect). It may be legitimate, therefore, to assess endpoints that would not be considered a basis for an RfD, but nonetheless affect quality of life. For example, avoidance of headaches and dizziness from perc inhalation may be validly considered as benefits of regulation of workplace levels, even if they are not strictly "toxic" effects.

Third, since severity can vary a good deal, it becomes especially important to identify the nature of the toxic effects that may be engendered. As with cancer assessment, traditional methods do not specify what effects may be risked at doses above those deemed "safe," and it is not generally presumed that humans will have the same toxic effects as those seen in experimental animals, but such presumptions are necessary for benefits analysis to gain specificity.

Fourth, unlike cancer, the severity (and not just the frequency) of response increases with increasing dose. Much toxicity data is expressed in quantal form (with or without an effect of a given grade), and the increasing health impact of higher doses on those individuals showing effects may not be readily described. Since people vary in their tolerance of exposures to agents, at some doses, some individuals will respond and others will not. At higher doses, more people in an exposed population will respond, but those who already responded at a lower concentration will have more severe effects at a higher one. As a consequence, the mix of severity of responses will vary with dose.

Fifth, unlike cancer, which once started becomes autonomous and independent of the dose that caused it, noncancer effects may (or may not) be dependent on continued exposure. For exposures that can be avoided, and for toxic effects that become evident with relatively short latency, it may be that sufferers of moderate symptoms remove themselves from exposure and limit the impact on their health. (Of course, the need to do so might be considered a non-health impact to which value might be ascribed.) For example, someone experiencing mild neurological symptoms from perc exposure on the job might seek reassignment. On the other hand, an effect on a pregnancy outcome provides no opportunity to detect and avoid a developing problem.

Sixth, in a similar vein, different endpoints will have different latencies, typical durations, tendencies to progress or resolve, and different degrees of recovery or reversibility being possible. The impact on quality of life will depend heavily on whether the effect appears early or late in life, whether it is permanent or reversible, and whether it gets worse with time, with or without continued exposure. These are not matters treated in traditional assessments of noncancer risk.

Seventh, many noncancer endpoints are defined and measured in terms of markers or indicators of effects, but the endpoints themselves are not the primary concern. For example, the effects of perchloroethylene on finger-tapping frequency or color discrimination are examined because these objective tests are thought to be measurable manifestations of underlying neurological impacts. The benefit of restricted exposure is not in better finger-tapping ability or fineness of color discrimination, but in freedom from the underlying neurotoxicity that these markers are presumed to reflect. The quantitative connection of marker effects with the impairments of the underlying system being affected are not always very clear.

Eighth, the reason that differing levels of response are seen at different doses for effects presumed to have a threshold is that different individuals have different tolerances, or individual thresholds, or degrees of reserve capacity. Those who respond at the lowest doses will be those in the population with the least reserve capacity. It may be that such people are very nonrandomly distributed over age and other demographic categories, and it may be that those prone to response are prone because of pre-existing ill health or marginal health, and their change in health state may be different than is assumed if effects are thought to fall randomly on the exposed members of a population.

Finally, traditional approaches to noncancer risk assessment make little attempt to characterize the quantitative changes in probability of response with changing dose levels. The focus is on finding NOAELs or benchmark doses—doses substantially without effect—rather than to map the shape of the dose-response relationship. Moreover, the means to extrapolate effects from animals to humans are not as well developed as for cancer assessment. The extrapolations are covered by "uncertainty factors" that act in part to make extrapolation corrections (to human equivalent doses or to particularly sensitive humans) and in part to allow for case-by-case uncertainty about how big an extrapolation correction to make. That is, the analysis is more of a safety assessment than a risk assessment, and impacts of exposures above the RfD are not readily characterized.

Methodological changes are needed that make noncancer risk analysis capable of explicit estimation and extrapolation. This requires separating the two roles of the uncertainty factors into unbiased estimates of extrapolation and additional allowances for uncertainty in those extrapolations. One promising approach is to use (in place of fixed uncertainty factors) empirical distributions over many agents of the magnitude of extrapolation needed. Baird *et al*. (1996) have explored such an approach. In current ongoing work, I and my colleagues (Sandra Baird, John Evans, Paige Williams, Andrew Wilson) are further developing this approach for the assessment of reproductive and developmental toxicity of ethylene oxide in humans. We use empirical distributions over many chemicals of species differences in toxicologically equivalent doses for noncancer effects as well as empirical information about interindividual variation in sensitivity to arrive at unbiased estimates (with characterization of uncertainty) of the human dose-response relationship. Such results are suitable for making estimates of impacts of exposures at different dose levels, including those above traditionally defined reference doses. The result of this analysis is a set of distributions of the uncertainty in doses expected to lead to different levels of response in an exposed human population, the kind of assessment that is needed for analysis of benefits of regulation for noncancer endpoints.

5 Conclusions

Risk analysis in support of benefits assessment is different in aims from analysis for the setting of regulatory levels as currently practiced. It needs to be focused on estimation of effects, not on the bounding of regions of exposure where one can be very confident that unacceptable impacts are not to be expected. Accordingly, methods of risk analysis for benefits assessment need to be somewhat different.

There are many profound challenges, but I have tried to show that they are approachable, at least in concept. The appropriate analyses are not quick or easy, and there is no minor tweak to existing methods that will make them fully applicable. Having laid out attempts to define the ideal analysis, perhaps simpler versions that are more readily conducted will become evident.

It is important to distinguish the task of estimating actual health effects (and the uncertainty about that estimation) from the task of setting regulatory levels. The difficulty in estimating benefits should be clear from the above discussion. A good deal of judgment is necessary, and there is likely to be controversy in specific cases about the weights to be put on alternative possible estimates of the health effects engendered by an exposure. This makes it difficult to use analysis of benefits and costs to define what acceptable exposure levels should be. This being said, there is value in using such analysis to gauge how much value is gained from regulation, and how much uncertainty there is about the magnitude of such gain.

Acknowledgements

Leslie Beyer and Eric Dubé of Gradient Corporation contributed to the summary of toxicity of perchloroethylene. Production of this paper was supported by the U.S. Environmental Protection Agency's Office of Air and Radiation under Order No. 0D-6263-NALX.

References

ATSDR (Agency for Toxic Substances and Disease Registry) 1997. Toxicology Profile for Tetrachloroethylene: Update. US Department of Health and Human Services, Public Health Service, Atlanta, GA.

Baird SJS, Cohen JT, Graham JD, Shlyakhter AI, Evans JS. 1996. Noncancer risk assessment: a probabalistic alternative to current practice. Human Ecol. Risk Assessment 2:79-102.

EPA (US Environmental Protection Agency) 1985. Health Assessment Document for Tetrachloroethylene (Perchloroethylene). (EPA/600/8-82/005FA) US Environmental Protection Agency, Office of Health and Environmental Assessment, Washington, DC.

EPA (US Environmental Protection Agency) 1991. Response to Issues and Data Submissions on the Carcinogenicity of Tetrachloroethylene (Perchloroethylene). Review Draft. (EPA/600/6-91/002A) US Environmental Protection Agency, Office of Research and Development, Washington, DC.

Evans JS, Gray GM, Sielken RI Jr., Smith AE, Valdez-Flores C, Graham JD. 1994. Use of probabalistic expert judgment in uncertainty analysis of carcinogenic potency. Regul. Toxicol. Pharmacol. 20:15-36.

IARC (International Agency for Research on Cancer) 1995. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Vol.63. Dry Cleaning, Some Chlorinated Solvents and Other Industrial Chemicals. World Health Organization, International Agency for Research on Cancer, Lyon, France.

ILSI (International Life Sciences Institute) 1998. Aggregate Exposure Assessment: An ILSI Risk Science Institute Workshop Report. Risk Science Institute, International Life Sciences Institute, Washington, DC.

Kaplan S, Garrick BJ. 1981. On the quantitative definition of risk. Risk Anal. 1:11-27.

McKone TE, Bogen KT. 1992. Uncertainties in health-risk assessment: an integrated case study based on tetrachloroethylene in California groundwater. Regul. Toxicol. Pharmacol. 15:86-103.

NCI (National Cancer Institute) 1977. Bioassay of Tetrachloroethylene for Possible Carcinogenicity. Publication No. 77-813. National Cancer Institute, Department of Health and Human Services. Bethesda, MD.

NTP (National Toxicology Program) 1986. Toxicology and carcinogenesis studies of tetrachloroethylene (perchloroethylene) (CAS No. 127-18-4) in F344/N rats and B6C3F$_1$ Mice (Inhalation Studies). National Toxicology Program Technical Report TRY 311. National Institutes of Health.

OEHHA (Office of Environmental Health Hazard Assessment) 1999. Public Health Goal for Tetrachloroethylene in Drinking Water. Review Draft. Pesticide and Environmental Toxicology Section, Office of Environmental Health Hazard Assessment, California Environmental Protection Agency.

Thompson KM, Evans JS. 1997. The value of improved national exposure information for perchloroethylene (perc): a case study for dry cleaners. Risk Anal. 17:253-271.

Appendix F-4
Calculating the Economic Benefits of Reductions in Manganese Air Concentrations
White Paper by Dr. Bernard Weiss,
University of Rochester, Rochester, NY

Assessing Benefits of Reductions in Manganese Air Concentrations

Bernard Weiss

Department of Environmental Medicine
University of Rochester
School of Medicine and Dentistry

Dr. Bernard Weiss
University of Rochester Medical Center
Rochester NY 14642

Tel: 716-275-1736
Fax: 716-256-2591
e-mail: bernard_weiss@urmc.rochester.edu

Introduction

Manganese presents a conundrum for risk assessment because it is both an essential nutrient and a potent neurotoxicant. Its neurotoxic properties have emerged almost exclusively from inhalation exposures, although some epidemiological data suggest that high concentrations in drinking water may be associated with neurological impairment. Several kinds of occupations expose workers to inhaled manganese, the most prominent being mining, ore-crushing, and ferro-manganese production. Mining for manganese ore provides the best documented association owing to the high levels of MnO_2 dust encountered in the process.

Table 1 lists some of the characteristic signs and symptoms of manganese neurotoxicity. Some, like bradykinesia, are also distinguishing signs of Parkinson's disease. Others, like the kind of emotional lability marked by abnormal laughing (and crying), are distinctive for manganese. In South American mining communities familiar with manganese intoxication, such a syndrome has earned the label, "locura manganica," or manganese madness, often viewed as the first stage of the full syndrome of manganese intoxication.

The most suitable animal model for research into manganese neurotoxicity is the nonhuman primate. Because of the unique organization of the primate brain, other animal models, such as rodents, are not as satisfactory, although they may yield useful information about neurochemical processes. Figure 1 depicts these differences as the relationship between dose and measure and shows, roughly speaking, a difference in sensitivity of close to two orders of magnitude between primates and rodents. One factor that may account for some of the discrepancy is the lack of advanced tests for motor function in the rodent studies comparable to the effortful response criterion used by Newland and Weiss (1992) in trained monkeys.

Although the motor signs exhibited by Mn miners correspond in part to those seen in Parkinson's disease (PD), enough differences are visible to question the widely-held proposition that Parkinson's and manganism are virtually identical. Barbeau (1984) suggested that the syndrome more closely resembled a dystonia than classical PD, a point of view also supported by Pat et al (1999) and others. Neuropathological observations support this distinction. With manganism, the main evidence of degeneration is seen in the globus pallidus, with less severe damage in the striatum (putamen and caudate nucleus) and in the substantia nigra pars reticulata. In contrast, the primary lesions seen in PD lie in the substantia nigra pars compacta and consist of depigmented and missing neurons, viewed as the dominant morphological markers of PD, accompanied by Lewy bodies, which consist of abnormally aggregated proteins found largely in dopaminergic neurons and recently shown to also contain the protein alpha-synuclein.

Convincing evidence of globus pallidus involvement also comes from magnetic resonance imaging (MRI) data. Because manganese is a paramagnetic metal, it modifies the return of protons to their original orientation after displacement by a strong magnetic field. These shortened times can then be used to produce different degrees of brightness in the calculated image that are related to local manganese concentration. The images and plots published by Newland et al (1989) and Newland and Weiss (1992) show the highest concentrations in the vicinity of the globus pallidus in exposed monkeys. MR images of an arc welder who had been exposed in the process of repairing and recycling railroad track made of manganese steel alloy also showed localized deposition in the globus pallidus (Nelson et al, 1993).

Ingested manganese is closely regulated by the gut. Inhaled manganese bypasses the gut, and can enter the brain in two ways. First, as described by Tjalve and Henriksson (1999), the olfactory pathways provide a direct path into brain tissue. Rats given ^{54}Mn intranasally accumulated the metal in a variety of brain structures, including the basal ganglia. Primates exposed by inhalation to trace amounts of ^{54}Mn showed a rise in brain levels that peaked at about 40 days (Newland et al, 1987). The manganese disappeared from brain much more slowly, with half-lives of 223 to 267 days in the two monkeys studied. ^{54}Mn was detected in the lungs for 500 days after exposure, suggesting that they served as a reservoir for uptake into brain (Figure 2). Although these data may also reflect some storage in bone, as noted by Andersen et al (1999), they indicate the strong possibility that long residence times in the lung provide a continuing source of brain exposure. This may be a special problem for young children in areas where dense vehicular traffic deposits manganese-laden dust. As with lead (cf., Lanphear et al, 1998), typical children's activities in high dust areas will expose them to elevated levels of both inhaled and ingested manganese, and Dorman et al (2000) have recently shown that neonatal rats administered manganese orally may be at greater risk for Mn-induced neurotoxicity than adult rats.

Most of the data pertaining directly to the benefits issue come from occupational studies. Table 2 gives the details of some of the important studies that attempted to relate exposure to neurobehavioral endpoints. The mean blood concentrations of exposed workers, except for Chia et al (1993), hover near 10 ：g/L, with their controls at one-half to two-thirds that value. Chia et al, however, studied a population in Singapore whose dietary habits undoubtedly differed from those in the other studies. Table 3 compares the results of a number of studies based on neurobehavioral endpoints. What is most evident there is the apparent sensitivity of motor function tests to manganese exposure, a result consistent with the evidence showing the main sites of deposition to lie in the basal ganglia, particularly the globus pallidus. Table 4 (Lucchini et al, 1999) offers more recent data from the population studied by Lucchini et al (1995). It too shows that mean control blood values are two-thirds of exposed values, meaning that an elevation

of one-third above baseline accounts for the performance differences between the two populations of workers. Also, note the closely overlapping ranges. Figure 3 plots the relationship in this population between air concentration and blood level in the work environment. Two features deserve comment. One is that even negligible air concentrations are associated with blood levels, as noted above, not overwhelmingly different from much higher concentrations. The second is that, at least in these workplaces, the distribution of exposure, as the authors note, is log-normal, with most workers clustered at the low end.

Worker populations present special problems for risk assessment. The healthy worker effect, a notorious confounder in epidemiological investigations, reduces the accuracy with which occupational data can be extrapolated to groups such as children, the elderly, or other especially susceptible populations. Moreover, standards such Threshold Limit Values and Permissible Exposure Limits are based on 8-hour days and 40-hour weeks rather than continuous environmental exposure. To more directly determine potential manganese toxicity in the general population, Mergler et al (1999) undertook a community study in southwest Quebec. The subjects ranged from 20 to 69 years of age and had not experienced any workplace exposures. The entire study sample of 297 subjects was about equally divided between men and women.

Table 5 presents the blood values. They show slightly higher levels in the women than in the men, but totally overlapping ranges. The investigators administered the most extensive series of neurobehavioral tests ever used to study manganese, and based most of their analyses on a separation of subjects based on blood levels. A value of 7.5 ：g/L served as the dividing concentration. Age was chosen as a second dichotomous variable separating subjects below and above 50 years of age.

The neuropsychological measures adopted by Mergler et al (1999) and that documented evidence of adverse effects are listed in Table 6. The first four are indices of motor function and the first three are described at length by Beuter et al (1999). The motor function measures yielded convincing relationships, but their most interesting features are their dependence on age. Figure 4 displays the interaction between manganese blood level (above or below 7.5 ：g/L) and age (above or below 50 years) for the index used to describe performance on a task requiring the subject to alternate strikes with a stylus at two spatially separated targets. This pattern, showing a persuasive influence of age, is consistent with most of the data from this study.

Neurodegenerative diseases, like most other degenerative diseases, are typically diseases of aging, with both incidence and prevalence rising with advancing age. One useful way to contemplate the potential impact of neurotoxic chemicals is to evaluate how they might shift the relationship between prevalence and age. A model of how even small shifts in a population distribution

can incur large public health costs is seen in Figure 6. It depicts the consequences of a 3-point or 3% shift in mean IQ score, the kind of shift produced by small elevations in lead exposure. It shows that even that small a shift produces a significant increase in the number of individuals classified as mentally retarded. It incurs massive expenses in remedial care and education, but also produces a significant decrease in the number of individuals in the superior range (e.g., IQ>130). Figure 7 shows that even a 1% leftward shift, or one IQ point, is a significant societal burden. In its evaluation of the benefits stemming from the removal of lead from gasoline, EPA, basing its calculations on the relationship between IQ score and lifetime earnings, estimated benefits approximating one trillion dollars.

A variation of this logic can be applied to manganese given the assumption that it can contribute to neurodegenerative disease. First, consider Figure 8, which depicts the reduction in nerve cell density with age that occur in certain brain structures. McGeer et al (1988) plotted the relationship between age and nerve cell number in the substantia nigra (SN). A key pathological marker of PD is loss of pigmented neurons in one part of SN. Figure 8 demonstrates that an acceleration of this natural loss by 0.1% annually will, over several decades, produce what might be termed premature aging of this structure. If the natural course of aging produces a loss of 40% by age 73, say, an additional acceleration of 0.1% will incur such a loss about ten years earlier.

Assume exposure to an agent that produces such a superficially minor acceleration. Figure 9 shows the consequences for the prevalence of PD of accelerations of 5 and 10 years respectively. The consequences are hardly minor. Table 7 takes the prevalence figures on which Figure 9 is based, and, from the projected US age distribution (US Census) in 2005, shows the baseline rates of PD and their estimated medical costs, and compares them to what would be expected if the age distribution were to be shifted by five years. The differences are considerable, and would result from an acceleration of functional loss of less than 0.1% annually (see Figure 8). For the age group 60-64, the increment in annual costs is over 700 million dollars.

Would this be a reasonable model for manganese? Or, put another way, what evidence is there to support a contribution by manganese exposure to PD or other neurodegenerative diseases?

One source of evidence is manganese poisoning, which confirms that manganese is a powerful neurotoxicant, producing the kinds of clinical signs, largely irreversible, listed in Table 1. A second source of evidence comes from detailed studies both of communities and of workers indicating that exposed populations displaying no signs of clinical disease can nevertheless be shown to suffer from neuropsychological deficits detected by appropriate testing procedures. But this kind of evidence is not specific to PD.

What is specific to PD, however, is both research and experimental data implicating the central nervous system structures targeted by manganese in PD. To incorporate these results into a benefits analysis first requires some probing into the potential relationship between manganese and neurodegenerative disease. It will be especially illuminating to examine how it might relate to PD because it is a clear example of a relationship with age. As noted earlier, the globus pallidus, on the basis of both chemical analyses and MRI, appears to accumulate manganese in greater quantities than other basal ganglia structures and is the site of lesions produced by manganese in appropriate doses. Although neuropathology does not indicate manganese-induced damage to the structure directly implicated in PD, the pars compacta of the substantia nigra, a great deal of evidence links its function with the globus pallidus.

One measure of the critical role played by GP in PD is the burgeoning literature on attenuation of PD symptoms by pallidal surgery or stimulation. Electrical stimulation of the internal pallidum may reduce the fluctuations associated with medication such as L-dopa, and permit a reduction in dosage. Pallidal surgery is now an accepted method for bringing substantial relief to PD patients. In addition, electrophysiological studies indicate a role for the globus pallidus in the resting tremor displayed by PD patients. Figure 5 shows the linkages among various basal ganglia structures and emphasizes the lack of isolation among them.

One conclusion to be drawn from this information is that what are called extrapyramidal diseases possess commonalities arising from their intimate and extensive structural and chemical interconnections. Damage to one component of the basal ganglia almost surely is bound to exert influence on functions subserved by other structural components, as in the overlapping symptoms of PD and Alzheimer's disease. In addition, the disabling effects of pharmacological therapies for PD, such as the dyskinesias resulting from L-dopa, are improved by pallidal stimulation, another source of evidence for the intimate links between GP and SN. A neat piece of evidence come from an experiment with monkeys (Zhang et al, 1999) rendered hemi-parkinsonian by an injection into the right carotid artery of MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine), originally discovered as a contaminant in designer drugs that produced parkinsonian signs in drug addicts. Figure 10 shows that monkeys with pallidal damage resulting from MPTP were less responsive to the amelioration of PD signs than monkeys lacking evidence of damage.

Both community surveys (Mergler et al, 1999) and studies of worker populations (e.g., Apostoli et al, 2000) suggest that relatively small increments in manganese blood levels are associated with significant diminutions in neurobehavioral function. If these functional indices are assumed to reflect deficits in brain function, and if we pair these deficits with the recognized declines in brain compensatory capacity associated with aging, slight elevations in

airborne manganese might produce a small, but medically and economically significant shift to an earlier onset of neurodegenerative diseases such as Parkinson's disease.

" Small" and "significant" need to be seen in context. An aging population is beginning to confront us with difficult medical and economic choices, and the most overwhelming problem is certain to be neurodegenerative diseases. In evaluating the potential contributions of environmental neurotoxicants to this problem, a simple calculation will prove illuminating. If the entrance of 30 patients into institutional care is delayed by one year, the savings amount to over one million dollars.

Reference List

Andersen ME, Gearhart JM, Clewell HJ . Pharmacokinetic data needs to support risk assessments for inhaled and ingested manganese. Neurotoxicology 1999;20:161-71.

Apostoli P, Lucchini R, Alessio L . Are current biomarkers suitable for the assessment of manganese exposure in individual workers? Am J Ind Med 2000;37:283-90.

Barbeau A . Manganese and extrapyramidal disorders (a critical review and tribute to Dr. George C. Cotzias). Neurotoxicology 1984;5:13-35.

Beuter A, Edwards R, deGeoffroy A, Mergler D, Hudnell K . Quantification of neuromotor function for detection of the effects of manganese. Neurotoxicology 1999;20:355-66.

Chia SE, Foo SC, Gan SL, Jeyaratnam J, Tian CS . Neurobehavioral functions among workers exposed to manganese ore. Scand J Work Environ Health 1993;19:264-70.

Dodel RC, Singer M, Kohne-Volland R, Szucs T, Rathay B, Scholz E, Oertel WH. The economic impact of Parkinson's disease. An estimation based on a 3-month prospective analysis. Pharmacoeconomics 1998;14:299-33312.

Dorman DC, Struve MF, Vitarella D, Byerly FL, Goetz J, Miller R . Neurotoxicity of manganese chloride in neonatal and adult CD rats following subchronic (21-day) high-dose oral exposure. J Appl Toxicol 2000;20:179-87.

Durif F, Lemaire JJ, Debilly B, Dordain G. Acute and chronic effects of anteromedial globus pallidus stimulation in Parkinson's disease. J Neurol Neurosurg Psychiatry 1999;67:315-22.

Fearnley JM, Lees AJ. Ageing and Parkinson's disease: substantia nigra regional selectivity. Brain 1991;114:2283-301.

Fine J, Duff J, Chen R, Chir B, Hutchison W, Lozano AM, Lang AE. Long-term follow-up of unilateral pallidotomy in advanced Parkinson's disease. N Engl J Med 2000;342:1708-14.

Goyer R. Toxic effects of metals. Amdur MO; Doull J; Klaassen CD, eds. Casarett and Doull's Toxicology. 4th ed. New York: Pergamon Press; 1991. pp. 623-80.

Gross RE, Lozano AM . Advances in neurostimulation for movement disorders.

Neurol Res 2000;22:247-58.
Hauber W, Lutz S . Dopamine D1 or D2 receptor blockade in the globus pallidus produces akinesia in the rat. Behav Brain Res 1999;106:143-50.

Hochberg F, Miller G, Valenzuela R, McNelis S, Crump KS, Covington T, Valdivia G, Hochberg B, Trustman JW. Late motor deficits of Chilean manganese miners: a blinded control study. Neurology 1996;47:788-95.
Huang CC, Chu NS, Lu CS, Chen RS, Calne DB . Long-term progression in chronic manganism: ten years of follow-up. Neurology 1998;50:698-700.
Kessler II. Epidemiologic studies of Parkinson's disease. 3. A community-based survey. Am J Epidemiol 1972;96:242-54.
Kim Y, Kim KS, Yang JS, Park IJ, Kim E, Jin Y, Kwon KR, Chang KH, Kim JW, ParkSH, et al. Increase in signal intensities on T1-weighted magnetic resonance images in asymptomatic manganese-exposed workers. Neurotoxicology 1999;20:901-7.
Lanphear BP, Matte TD, Rogers J, Clickner RP, Dietz B, Bornschein RL, Succop P, Mahaffey KR, Dixon S, Galke W, et al. The contribution of lead-contaminated house dust and residential soil to children's blood lead levels. A pooled analysis of 12 epidemiologic studies. Environ Res 1998;79:51-68.
Lombardi WJ, Gross RE, Trepanier LL, Lang AE, Lozano AM, Saint-Cyr JA . Relationship of lesion location to cognitive outcome following microelectrode-guided pallidotomy for Parkinson's disease: support for the existence of cognitive circuits in the human pallidum. Brain 2000;123(Pt 4):746-58.
Lozano AM, Lang AE, Hutchison WD. Pallidotomy for tremor. Mov Disord 1998;13(Suppl 3):107-10.
Lucchini R, Apostoli P, Perrone C, Placidi D, Albini E, Migliorati P, Mergler D, Sassine MP, Palmi S, Alessio L . Long-term exposure to "low levels" of manganese oxides and neurofunctional changes in ferroalloy workers. Neurotoxicology 1999;20:287-97.
Lucchini R, Selis L, Folli D, Apostoli P, Mutti A, Vanoni O, Iregren A, Alessio L . Neurobehavioral effects of manganese in workers from a ferroalloy plant after temporary cessation of exposure. Scand J Work Environ Health 1995;21:143-9.
McGeer PL, Itagaki S, Akiyama H, McGeer EG. Rate of cell death in Parkinsonism indicates active neuropathological process. Ann Neurol 1988;24:574-6.
Mergler D, Baldwin M, Belanger S, Larribe F, Beuter A, Bowler R, Panisset M, Edwards R, de Geoffroy A, Sassine MP, et al. Manganese neurotoxicity, a continuum of dysfunction: results from a community based study. Neurotoxicology 1999;20:327-42.
Nelson K, Golnick J, Korn T, Angle C . Manganese encephalopathy: utility of early magnetic resonance imaging. Br J Ind Med 1993;50:510-3.
Newland MC . Animal models of manganese's neurotoxicity. Neurotoxicology

1999;20:415-32.
Newland MC, Ceckler TL, Kordower JH, Weiss B . Visualizing manganese in the primate basal ganglia with magnetic resonance imaging. Exp Neurol 1989;106:251-8.
Newland MC, Cox C, Hamada R, Oberdorster G, Weiss B . The clearance of manganese chloride in the primate. Fundam Appl Toxicol 1987;9:314-28.
Newland MC, Weiss B . Persistent effects of manganese on effortful responding and their relationship to manganese accumulation in the primate globus pallidus. Toxicol Appl Pharmacol 1992;113:87-97.
Pal PK, Samii A, Calne DB . Manganese neurotoxicity: a review of clinical features, imaging and pathology. Neurotoxicology 1999 1999;20:227-38.
Papa SM, Desimone R, Fiorani M, Oldfield EH . Internal globus pallidus discharge is nearly suppressed during levodopa-induced dyskinesias. Ann Neurol 1999;46:732-8.
Shinotoh H, Snow BJ, Hewitt KA, Pate BD, Doudet D, Nugent R, Perl DP, Olanow W, Calne DB . MRI and PET studies of manganese-intoxicated monkeys. Neurology 1995;45:1199-204.
Tjalve H, Henriksson J . Uptake of metals in the brain via olfactory pathways. Neurotoxicology 1999;20:181-95.
Vitarella D, Wong BA, Moss OR, Dorman DC. Pharmacokinetics of inhaled manganese phosphate in male Sprague-Dawley rats following subacute (14-day) exposure. Toxicol Appl Pharmacol 2000;163:279-85.
Weiss B . Vulnerability of Children and the Developing Brain to Neurotoxic Hazards. Environ Health Perspect 2000;108(Suppl 3):375-81.
Zhang Z, Zhang M, Ai Y, Avison C, Gash DM. MPTP-Induced pallidal lesions in rhesus monkeys. Exp Neurol 1999;155:140-9.

Appendix F-4 Figures and Tables

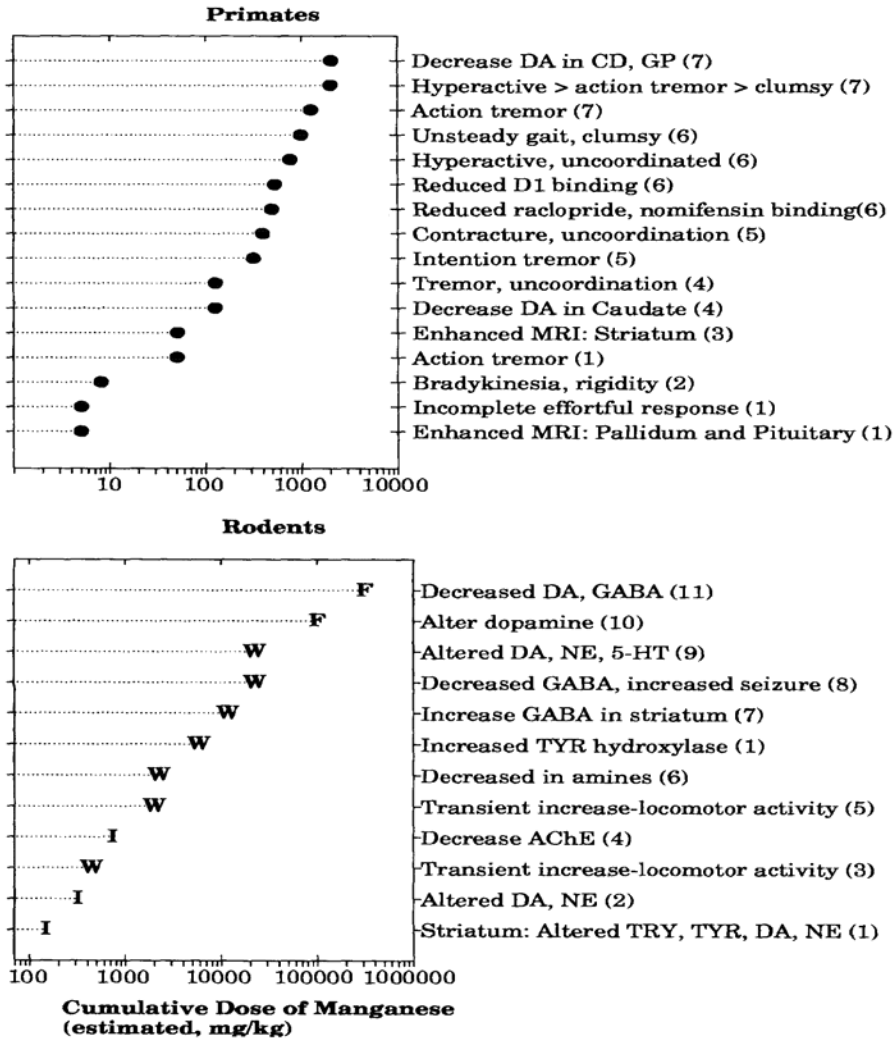


Figure 1. Relationships between administered manganese dose and indices of neurotoxicity in primates and rodents (Newland, 1999).

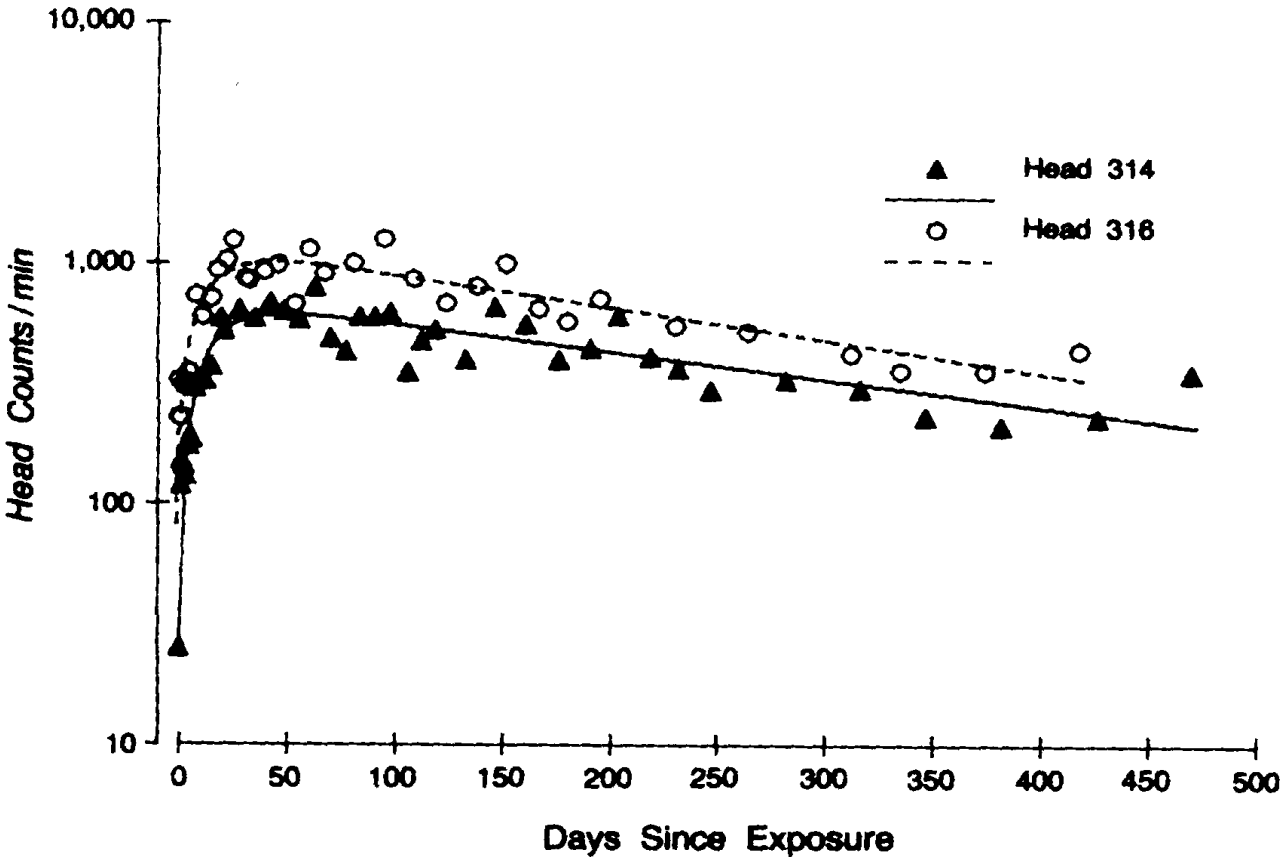


Figure 2. Radioactivity in the head after inhalation of ^{54}Mn in two monkeys (M. nemestrina). From Newland et al, 1987.

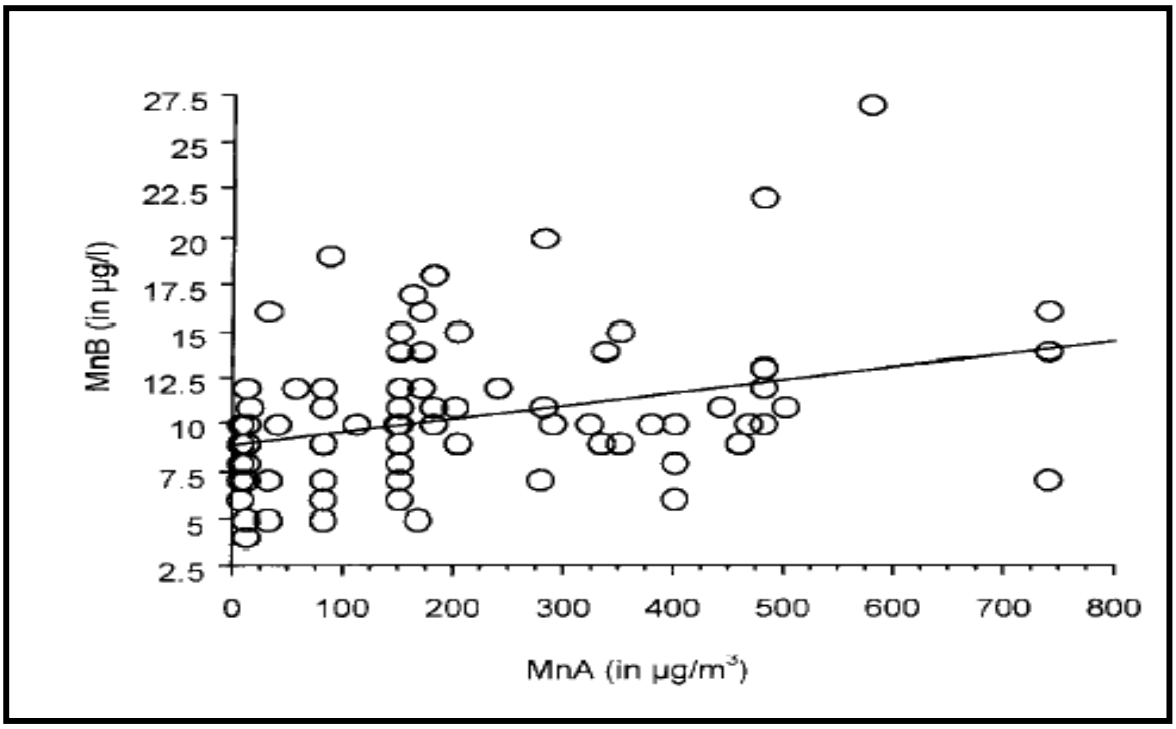


Figure 3. Relationship between air and blood manganese levels in exposed workers (Apostoli et al, 2000)

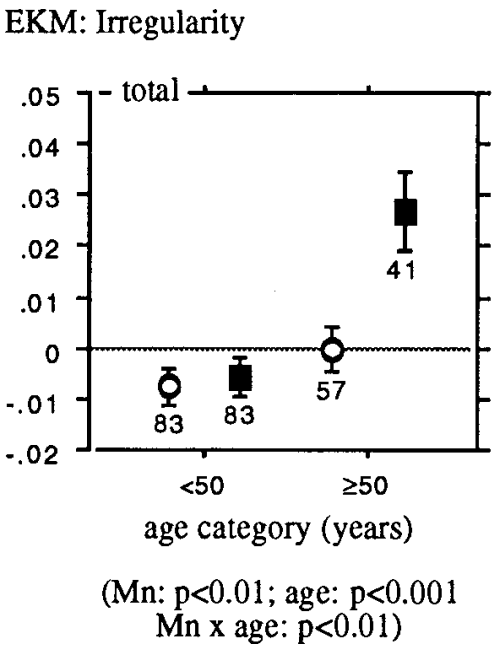


Figure 4. Performance on the Irregularity score of the task requiring alternate striking of spatially separated targets. Key: ■ = MnB > 7.5 µg/L. ○ = MnB < 7.5 µg/L. From Mergler et al, 1999.

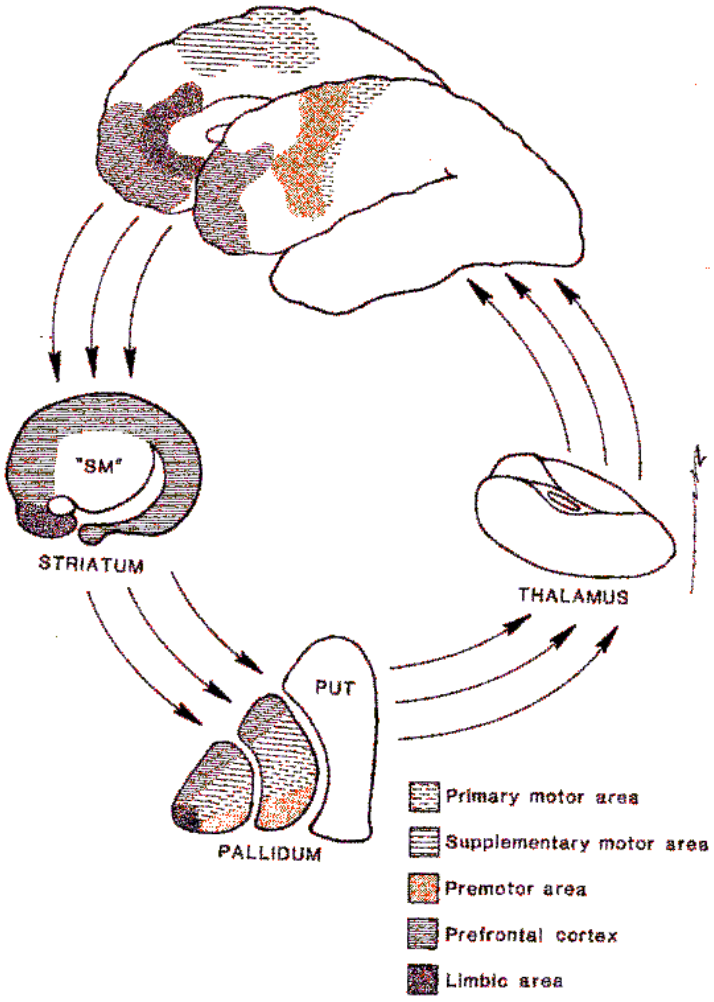


Figure 5. Globus Pallidus connections. PUT=putamen; SM=sensorimotor circuits.

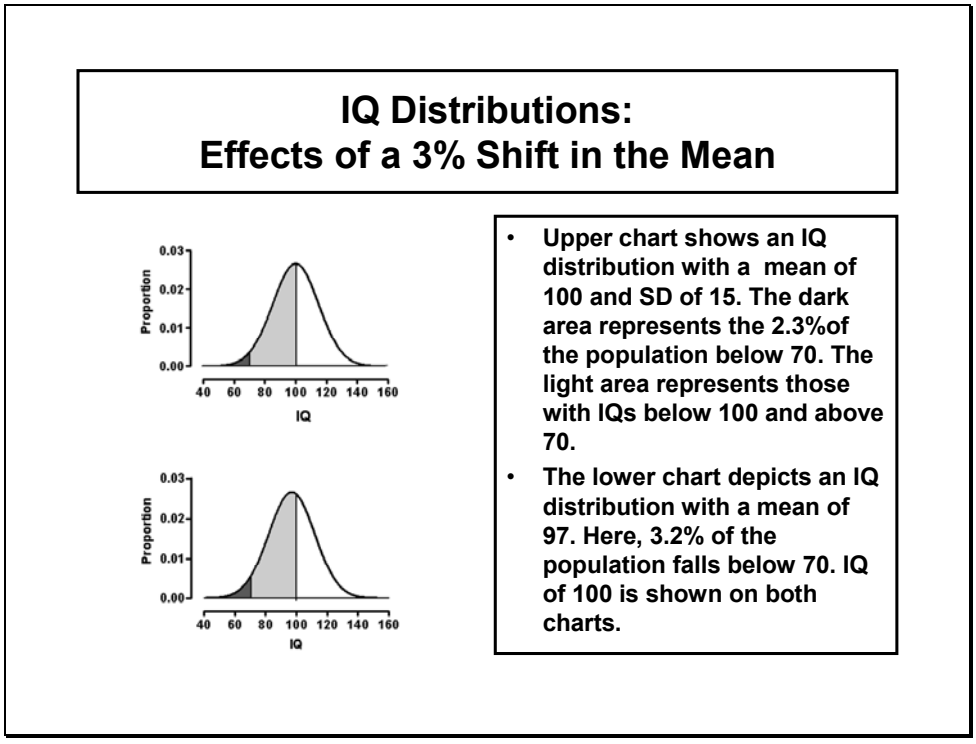
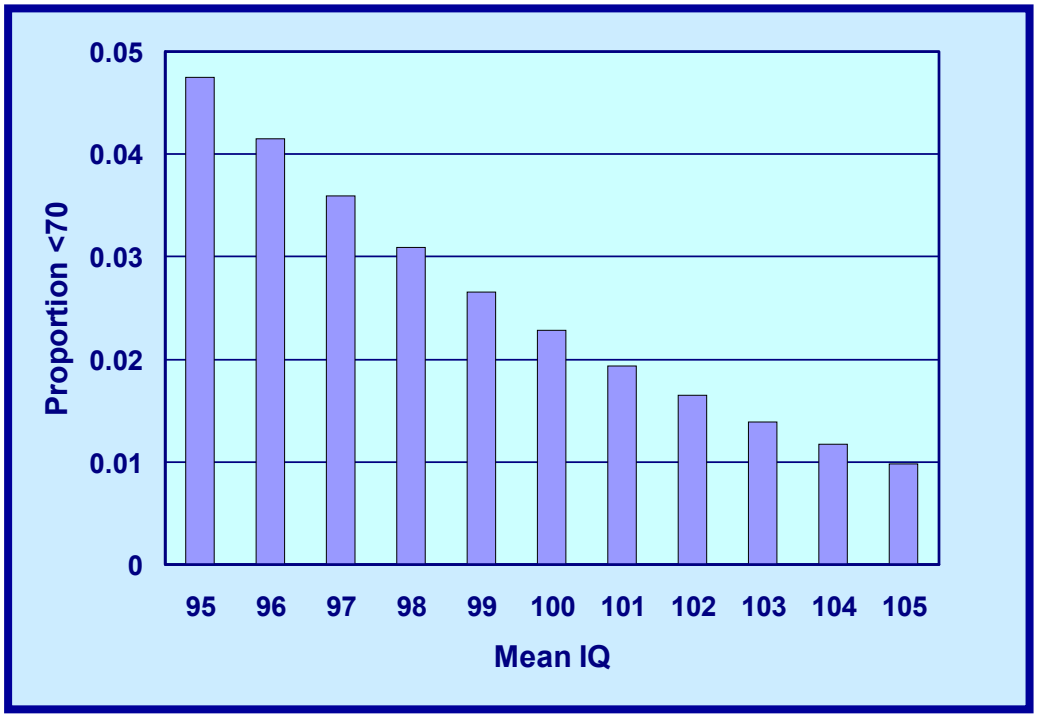


Figure 6. Consequences for classification of Mentally Retarded (IQ<70) of a 3% shift in the IQ distribution. From Weiss, 2000.

Figure 7. Proportion of Individuals in Retarded Range (IQ<70) with Different Population Mean IQ Scores



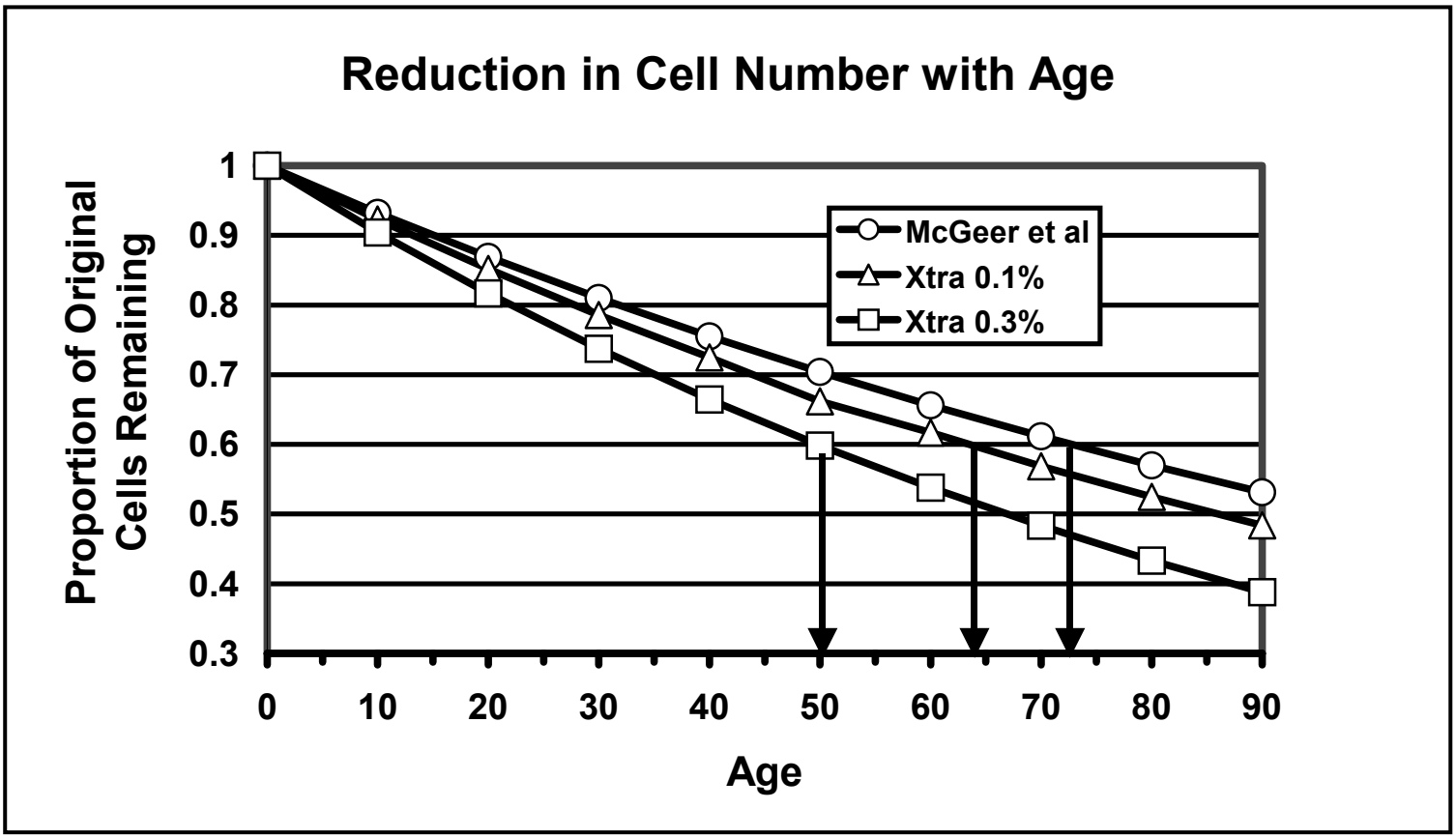


Figure 8. Cell loss with age in substantia nigra

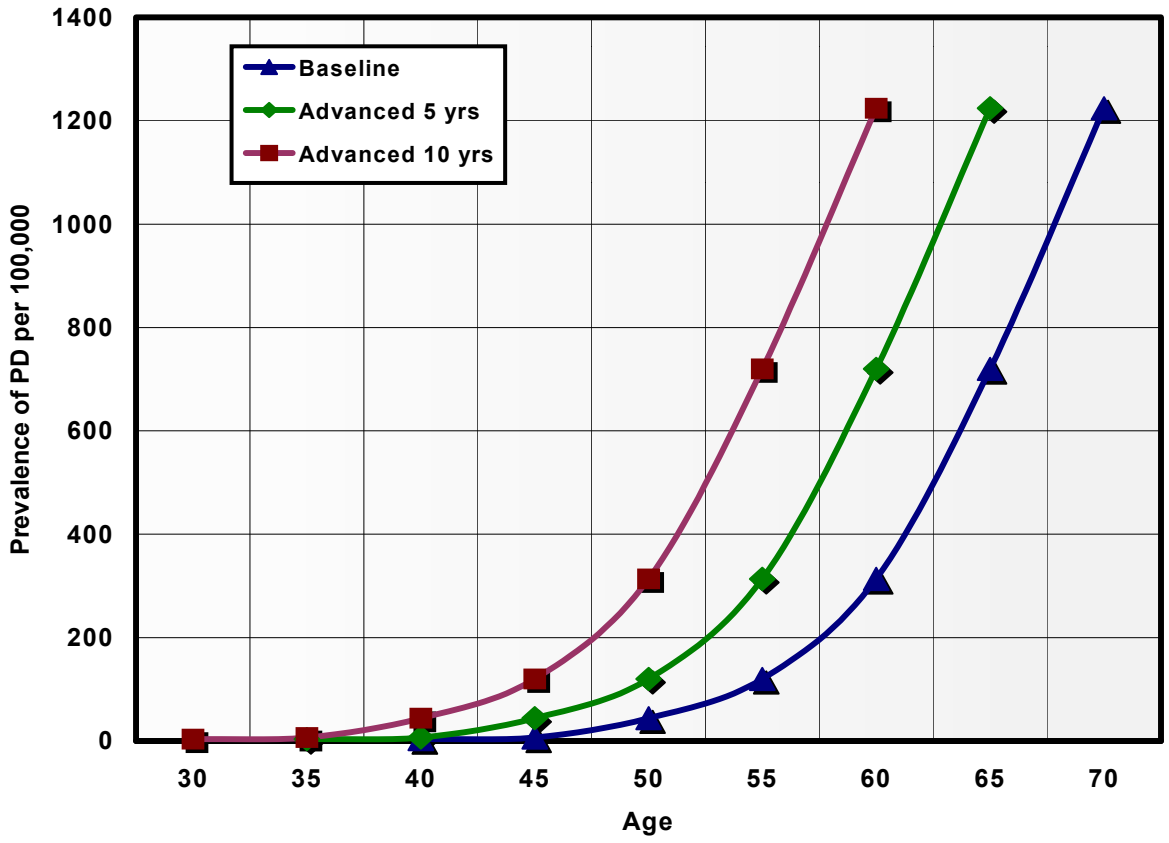


Figure 9. Acceleration of Parkinson's disease onset. From Weiss, 2000

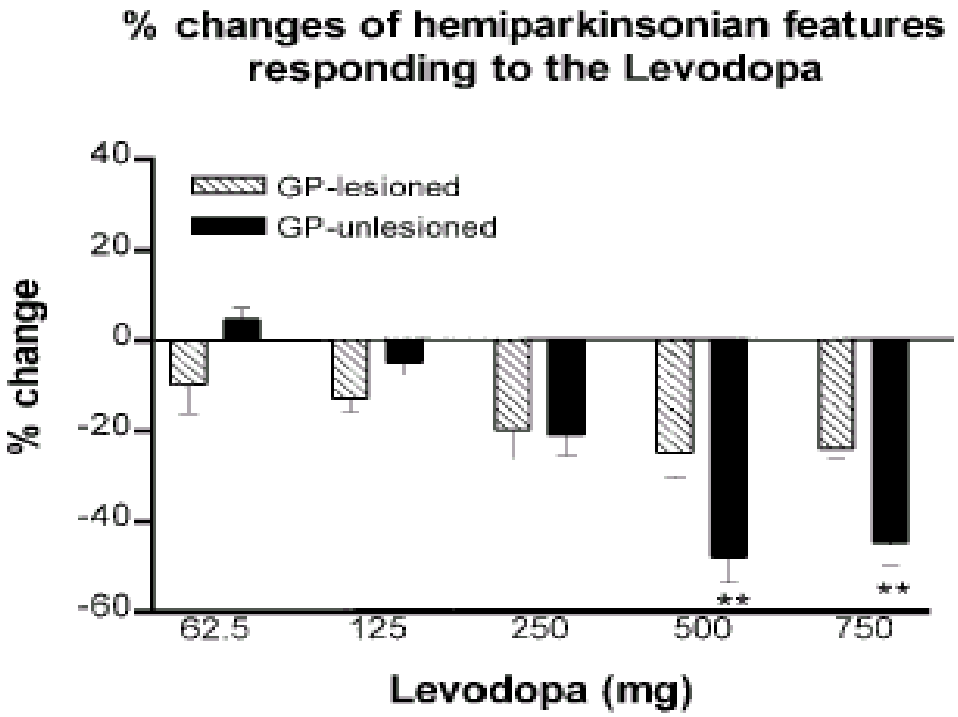


Figure 10. Comparison of therapeutic effectivenes of L-dopa in hemiparkinson monkeys with and without collateral pallidal damage.

Table 1. Signs and symptoms of manganese neurotoxicity

- Abnormal gait
- Impaired coordination
- Abnormal laughter
- Expressionless face
- Weakness
- Bradykinesia
- Somnolence
- Dysarthria
- Difficulty walking
- Clumsiness
- Lack of balance
- Muscle pains
- Diminished leg power

Table 2. (Mergler and Baldwin, 1997)

Demographics and Internal Exposure Parameters for a Number of Comparable Studies of Active Workers

Reference (exposed + controls)	Type of plant	Mean age of exposed	Years of exposure, mean ±SD (range)	MnB of exposed, geometric mean (μg/100 ml)	MnB of controls, geometric mean (μg/100 ml)	MnU of exposed, geometric mean (μ g/g cr)	MnU of controls, geometric mean (μg/g cr)
Roels *et al.* (1987) (141 + 104)	Mn oxide and salt	34.3 ± 9.6 (19–59)	7.1 ± 5.5 (1–19)	1.22*	0.49	1.59*	0.15
Wennberg *et al.* (1991) (30 + 90)	Steel smelting	46.4 (19–63)	9.9	—	—	—	—
Roels *et al.* (1992) (92 + 101)	Dry alkaline battery	31.3 ± 7.4 (22–49)	5.3 ± 3.5 (0.2–17.7)	0.81*	0.68	0.84*	0.09
Chia *et al.* (1993) (17 + 17)	Mn ore milling	36.6 ± 12.2	7.4 ± 4.3 (1–14)	2.53	2.33	6.1 μg/liter	3.9 μg/liter
Mergler *et al.* (1994) (74 + 74)	Ferro and silico alloy production	43.4 ± 5.4 (32–58)	16.7 ± 3.2 (1–17)	1.03*	0.68	0.73	0.62
Lucchini *et al.* (1995) (*n = 58*)[a]	Ferroalloy plant	38.9 (20–53)	(2–28)	1.19*	0.60	2.8 μg/liter	1.7 μg/liter

[a] Lucchini *et al.* compared 19 low-exposure workers to 39 more highly exposed workers.
* The authors report significant differences between those exposed and control ($P < 0.05$).

Table 3. (Mergler and Baldwin, 1997)

Results of Neurofunctional Assessment from Comparable Studies of Active Manganese-Exposed Workers

Reference (exposed + controls)	Motor functions	Hand steadiness	Response speed	Diadocho-kinesia	Memory	Other intellectual functions	Olfactory sensitivity	Mood state
Siegl and Bergert (1982) (25 + 21)	—	—	↓	—	—	—	—	—
Roels *et al.* (1987) (141 + 104)	↓	↓	↓	—	↓	—	—	—
Wennberg *et al.* (1991) (30 + 90)	↓	—	↓	—	↓	**n.s.**	—	—
Wennberg *et al.* (1992) (30 + 90)	—	—	—	↓	—	—	—	↓
Roels *et al.* (1992) (92 +101)	—	↓	↓	—	**n.s.**	—	—	—
Chia *et al.* (1993) (17 + 17)	↓	—	—	—	↓	↓	—	—
Mergler *et al.* (1994) (74 + 74)	↓	↓	**n.s.**	—	**n.s.**	↓	↑	↓
Beuter *et al.* (1994) (10 + 10)	—	—	—	↓	—	—	—	—
Lucchini *et al.* (1995) (n = 58)[a]	↓	—	**n.s.**	—	↓	↓	—	—

[a] Lucchini *et al.* compared 19 low-exposure workers to 39 more highly exposed workers.

Table 4. Blood manganese levels (µg/L) in exposed and control workers (Lucchini et al, 1999).

	Mean	Median	Range
Exposed	9.71	9.00	4-19
Control	6.00	6.00	2-9.5

Table 5. Blood manganese levels (µg/L) in the Quebec community study (Baldwin et al, 1999; Mergler et al, 1999)

	Mean	Median	Range
All	7.50		2.5-15.9
Women	7.90	7.70	2.8-15.9
Men	7.00	6.60	2.5-13.9

Table 6. Neuropsychological Measures used in the Quebec Community Study

- Eye-hand coordination
- Rapid pointing movements
- Tremor frequency and amplitude
- Neurological exam indices
- Learning and recall
- Mood
- Psychological symptoms
- Body sway

Table 7. Projected annual medical costs of Parkinson's disease, to age 74, based on age-related prevalence comparing baseline age distribution with onset accelerated by 5 years.

Age (2005)[1]	Number (x1,000)	Base PD	Base Cost (x$1,000)[2]	+5 PD	+5 Cost (x$1,000)	Difference (x$1,000)
35-39	**20,082**			**823**	**11,160**	**11,160**
40-44	**22,634**	**928**	**12,584**	**1,584**	**21,479**	**8,895**
45-49	**22,230**	**1,556**	**21,099**	**9,810**	**133,024**	**111,925**
50-54	**19,661**	**8,876**	**120,359**	**23,593**	**321,277**	**200,918**
55-59	**16,842**	**28,210**	**382,528**	**52,827**	**716,334**	**333,806**
60-64	**12,848**	**40,300**	**546,680**	**92,506**	**1,254,381**	**707,701**
65-69	**10,086**	**72,619**	**984,714**	**123,458**	**1,674,090**	**689,376**
70-74	**8,375**	**102,514**	**1,390,090**	**119,218**	**1,616,596**	**226,506**

[1]US Census Projections
[2] Based on Dodel et al (1998) @$13,560/yr

APPENDIX G: WRITTEN SUBMISSIONS FROM KEY DISCUSSANTS

Expert panelists and the Workshop Co-Moderator, Dr. Roy Albert, provided the following written comments after the workshop on the questions discussed on June 23, 2000:

G.1. Comments of Dr. Roy Alpert, Division of Environmental Health, University of Cincinnati

a) A Way to Characterize Overall Carcinogen Risk: If the EPA's Air Office is going to estimate the monetary benefits of regulating carcinogenc HAPs, how will it deal with agents whose probability of being a human carcinogen is less than certain? One can estimate the number of cancer deaths by use of a dose response function and an exposure estimate on the assumption that the agent in question is a human carcinogen. But suppose the evidence does not permit us to say that it is definitely a human carcinogen? What then? One possibility is to multiply the estimated number of cancer cases by a weighting factor that is determined by the strength of the evidence. Suggested values for weighting factors are given in the table below. The sue of the square of the weighting factors gives greater separation between strong and week evidence.

EPA Category	Descriptor	Weighting Factor	(Weighting Factor)^2	Level of Evidence
A	Definite	1.0	1	Sufficient human and anumal
B1	Probable	0.75	0.56	2 species postiive and some human
B2	Probable	0.5	0.25	2 species positive
C	Possible	0.25	0.06	1 species both sexes
C	Possible	0.12	0.01	1 species 1 sex

b) A Method for Best-Estimate and Uncertainty Characterization of Carcinogen Risks. The stated purpose of the meeting was the development of a berst-estimate and uncertaty characterization for hazard and dose-response functions fo use in benefit analysis of HAP regulations. One possibility is offered below. The proposed method avoids large downward extrapolations to vanishingly small levels of risk.

The EPA's approach to carcinogen risk estimation was developed by its Carcinogen Assessment Group (CAG) in the mid 1970s. The first method involved the use of the lowest statistically signfiicant data point as the point of departure for a linear non-threshold extrapolation to zero dose/zero incidence. Shortly after that, because of the emphasis on the conservative approach to risk assessment, the linear non-threshold extrapolation sued the 95% upper confidence limit of the lowest statistically significant data point as the point of deparrture for the extrapolation to zero dose/zero incidence. The risk estimates obtained this way were described as "plausible upper limit estimates, i.e., ones that were not likely to be highter than the true risks but could be lower even to a considerable extent." After a few years, the statisticians objected to the procedure, saying that all the data points above the lowest statistically significant point were being wasted. After much discussion, the use of the multistage model was then agreed upon and introduced by CAG and remains the method used up to the present.

The multistage model appealed to the statisticians because it is flexible enough to accommodate almost any data set and it has biological plausibility, since cancer is multistage in its development. Furthermore, the multistage model has a low-dose linear non-threshold component. This low-dose linear non-threshold component depends on the assumption that the carcinogen in question behaves like whatever it is that causes tumors in control animals (the background). The 95% upper confidence limit for the low-dose linear non-threshold extrapolation was carried over to the multistage model. In all the variants of the oinear extrapolation, the background tumor incidence was subtracted from the incidence obatained with each of the carcinogen doses so that the extrapolation could be extended to the zero dose/zero incidence. No consideration was taken of the statistical uncertainty in the background incidence.

A more realistic estimate of risk with confidence limits can be made while retaining the essential features of the established EPA approach. The multistage model can be retained on the basis of the original rationale for using it. The low-dose linear non-threshold component can also be retained on the grounds that carcinogens probably act the same way as the cuases of background tumors. However, in the proposed approach we do not subtract out the background tumor incidence in order to extrapolate down to zero dose/zero incidence. We extrapolate down to the background tumor incidence. This follows logically from the assumption that the carcinogen in question acts like whatever it is that causes background tumors. The dose response curve is readjusted so that the administered dose is an increment to the equivalent background dose as determined by the slope of the linear component of the multistage model. The risk can be expressed as either the absolute incremental risk or the relative incremental risk. For example a 1% absolute risk increment on a 5% background risk would be a relative incremental risk of 20%.

There are two statistical uncertainties. The first deals with how well the data fit the extrapolation model. The second uncertainty relates to the background cancer level. The overall uncertainty of the risk estimates would be the combination of the two uncertainties.

The above approach was published in the context of time to tumor data rather than life-time incidence. But the principle is the same. The reference is as follows:

Albert, RE. and B Altschuler, 1976. Assessment of Environmental Carcinogen Risks in Terms of Lifen Shortening. *Environmental Health Perspectives 13: 91-94.*

G.2. Comments of Dr. John C. Bailar III, Department of Health Studies, University of Chicago, Chicago, IL.

How best to identify limitations and uncertainties in both risk assessment methods and economic models.

Here I will summarize some of the important limitations and uncertainties, put them in a context, and offer some suggestions about how to deal with them.

Cost-benefit analysis is an information-hungry process, which we must apply to an information-sparse problem. This can be done, and it will be done, but the results will not be pretty.

Further, regulatory decision-making based on cost-benefit analysis is a precision-hungry process that we must necessarily base on precision-sparse inputs. Again, this can and will be done, but the results will not be pretty.

The two fundamental problems are the enormous burden of work required to deal with 189 HAPs and the great uncertainty inherent in the estimates on both sides of this process - the costs and the benefits -- for each one of them. It may be close to impossible for conscientious economists and conscientious risk assessors to do an even marginally competent job for each HAP. It may be even more difficult to remain honest about the real level of uncertainty. The more we have learned about any particular hazard, the more complex we have found it to be. This seems to be a general phenomenon, and we should adapt to it. There is no reason to think that any of the 189 HAPs is basically any simpler then benzene, though benzene looks quite complicated because we know a great deal about it.

We never think we know as much as the users of our analyses demand, and we never really know as much as we think we do. Uncertainty to three orders of magnitude is the norm in risk assessment. When that is compounded with the deep and numerous problems of cost-benefit analysis the uncertainties may very well rise to six orders of magnitude. This is in part because we need to estimate marginal effects, both costs and benefits, and these marginal effects are small differences in sometimes quite large basic numbers.

Some of the sources of uncertainty are:

a) Missing, incomplete, and inaccurate records of human exposure; for ambient HAPs, these gaps are enormous, and we know even less about probable levels of future exposures

b) A nearly complete absence of information about co-exposures

c) Poor and incomplete records of human health outcomes, including the whole range of biases recognized by epidemiologists (non-random dropouts, healthy worker effects, recall biases, changing concepts and tools for diagnosis of disease, and all the rest)

d) Unknown mechanisms of action, even when toxic endpoints have been identified, and the consequent difficulties of predicting what will happen at biologically low exposures

e) Targets that are moving over time, including the effects on future costs and benefits of increasing life span and increasing time in which delayed toxicities can appear, decreases in competing causes of illness and death, and the different and changing medical implications of various illnesses and impairments

f) The over-simplifications imposed by the regulatory context: Human disease rarely has one cause, and one cause does not always produce a single disease, so that we try to satisfy unifactorial regulatory demands in a highly multifactorial situation.

g) Numerous extrapolations, which (for animal studies) include: animal to human, high to low dose, one route of administration to another, constant lifetime to intermittent and/or irregular exposure, and uniform and protected laboratory environments to highly diverse and unprotected human situations

h) Large standard errors, sometimes from necessarily small samples, in many of the critical inputs

i) Unpredictable and poorly understood environmental transport, including meteorology, hydrology, and many other things

j) Possibly important synergies in biologic effects and a gross lack of understanding about two critical issues: what is in specific mixes of air pollutants, and how variable the mixes can be from one setting to another. Such information does not seem to be available for any subset of the HAPs.

k) Allocating the "blame" for bad outcomes that are in fact the result of synergies among exposures (e.g., when we cost out the extra risks and expenses from the synergy between asbestos and tobacco smoke, which industry should get the hypothetical bill?)

l) The non-linearity of many impairments; for example, one can lose fifty percent of liver capacity and never know it, so that loss of the first ten percent may have a value vastly different from the value of the last ten percent.

m) Unknown levels of regulatory compliance, and predictably incomplete efforts to

monitor compliance for 189 HAPs

n) Big and fundamental questions about which things we are to value and about how we are to attach specific values to those things

o) All the difficulties of trying to place dollar valuations on various kinds of incommensurate health outcomes, including death, and the even greater conceptual difficulties of placing values on markers of exposure or effect when signs or symptoms of illness have not appeared

p) The sheer volume of 189 HAPs, which will impose great demands for scarce technical talent as well as resources, and which will certainly force the adoption of means to keep those demands within reachable limits

q) Estimates of the costs of compliance with a new regulation are notoriously prone to error, usually but not always in the direction of gross over-estimation of what it will cost polluters to clean up their act

More fundamentally, different people will value things in different ways. Whose valuation counts? Will we take what people say the first time we ask them, or try to educate them before they give us their values? Will we give special weight to the valuations of people who have had the outcomes in question and understand them? What about substitution effects? What about benefits forgone? Will we assign the same value to every death or every illness of given severity? Willingness to pay is hardly a meaningful metric for someone who has barely enough to get by anyway.

Are dollars even the right metric? There are questions about equity, there are differing and non-linear utilities, there are major questions about what to exclude as externalities. There are discount rates and intergenerational effects to account for, as well as distributional effects more generally, and there may well be important transaction costs.

There are theoretical answers to all or nearly all of these points, but each application of theory requires the use of inputs that are to some extent uncertain. I was quite serious about the six orders of magnitude of uncertainty. We can be honest about that uncertainty, bury our heads in the sand, and see the special interests take over the process, or we can abandon our scientific and technical integrity, lie about the uncertainty, and ultimately lose our credibility and our claim to special standing as scientists. Or, we can come to grips with it as a serious challenge, deal with it directly and honestly, and do what we can to make sure that users of our analyses understand the fundamental falsehood of any claims (including tacit claims) that some other approach is better.

Some recommendations that may be constructive:

a) There is a need for substantial education about the art of the possible. We need to

educate congress, the public, and the news media; in fact, every group or person who encounters these issues. Perhaps most, we need to educate ourselves, so that we have realistic expectations of what we can in fact accomplish.

b) There is an evident need for very much closer links between risk analysts and cost-benefit analysts. Each person with a significant technical or managerial role on either side of this divide should spend at least six months working in the other program. (When I came to this point in the meeting itself, I heard snickers from the audience. As well as I could tell, they came from a few persons on each side of the divide, though there may be no better way to understand the real problems, and to learn how to help solve those problems, than to wrestle with them yourself. I fear that some persons may not recognize the career advantages of knowing both sides. Heavy pressure from higher levels may be needed to implement this change.)

c) I recommend also that there be regular, weekly meetings on each project in which risk analysis will be a significant element of a cost benefit analysis, to assure full communication and understanding about what is needed, what can be provided, and how to adjust for the inevitable gaps between these. Passing written reports back and forth will not do the job.

d) There is a need for serious attention to the level of accuracy needed at each step of the process of risk analysis / cost benefit analysis. Does it matter if we are off by 20%? Two-fold? Ten-fold? One thousand-fold? Analysts and managers rarely address these questions in any serious way (perhaps because they do not get beyond the correct recognition that greater accuracy is always better, ceteris paribus), and yet they are critically important here because of the need to make most effective use of limited resources and to balance countless compromises and trade-offs.

e) We need an organized, almost assembly-line approach to risk assessment if we are to deal with all 189 HAPs. The need for standardized procedures to deal with the HAPs inevitably leads to the need for "bundling", though this too will require some compromises, and general solutions may not always fit well. (Perhaps Procrustes had the right idea.) Bundling according to health endpoint or cause of death might advance the purposes of hazard identification. Bundling according to chemical species could advance exposure estimation. Similarly, dose-response and sensitivity studies might be stronger if we bundle by biologic mechanism or mode of action. Finally, regulatory considerations may fit best with bundling by source categories. The last of these is apparently favored by the economists because it is related directly to their task, but the value of the other axes of classification in the risk assessment phases may outweigh the value of using source categories in the cost-benefit analysis. It may be that we could somehow combine these other axes with source categories to gain some of the advantages of both.

f) The axis of bundling needs more study than has been evident here, since it has implications for the inputs to the cost-benefit analysis, which should follow the same

pattern. If the bundling is by source categories, testing should also be by source categories, with a focus on the study of the complex mix that comes from any one source, rather than its components. I understand the technical and scientific objections to this, but if we take those objections too seriously, they will simply undermine the whole rationale for bundling by source categories. One cannot have it both ways. These matters require serious study before EPA adopts an approach based on source categories as a critical axis of classification.

In summary, we need radical solutions; tinkering with methods, approaches, and mind-sets now on the shelf will not do the job. The technical and scientific issues of cost-benefit analysis of regulatory control of the 189 HAPs are daunting, and any conceivable result will inevitably carry an enormous margin of uncertainty. It is important that everyone involved in this process, including all users of the results, understand that this true and that it is unavoidable. This is not an indictment of either risk analysis or cost-benefit analysis, both of which are critically important in collecting, analyzing, and interrupting what is or will be on the record. It is certainly better to collect, organize, and interpret what we can than to simply give up and proceed on blind faith that some step is or is not justified. However, we must not expect to produce, or allow others to expect from us, a level of precision and certainty that the process is unable to deliver.

G.3. Economics and Toxicology: Results of a Dialogue on the Prospects for Assessment of Benefits from Regulation of Hazardous Air Pollutants. Comments of Dr. Trudy Cameron, Department of Economics, University of California, Los Angeles, CA.

For some time now, there has been a degree of acrimony between some economists and some toxicologists. This acrimony concerns the nature of the information being supplied by toxicologists to economists for use in the congressionally mandated task of valuing the non-market benefits of environmental regulations. From the economist's point of view, the problem can be characterized as "Why don't they just give us the information we need? Why are they being so uncooperative?" From the toxicologist's point of view, the problem can be characterized as "Why are they asking us for something that is impossible to provide? Why are they being so unreasonable." The recent EPA workshop should reduce this acrimony, and help us focus on the task at hand, by emphasizing the insight that we do not live in a perfect (research) world.

a) <u>What do economists need, in a perfect world, to calculate benefits?</u> People's demand for regulation of hazardous air pollutants is what economists call a "derived demand." People are frequently viewed as demanding environmental regulation not for its own sake, but mainly because the regulation may achieve a reduction in health risks.

Economists are accustomed to dealing with a wide range of derived demands. For example, few people demand electricity for its own sake. Instead, we are willing to pay for electricity because of the services that can be provided by electrical appliances. To understand willingness to pay in a derived demand context, it is helpful to consider the chain of relationships

that form the connection between willingness to pay and the underlying good. To continue the electricity example, consider somebody who uses electricity to heat water. Formal modeling of willingness to pay for a kWh of electricity for water heating will depend on the individual's value of a gallon of hot water and on the efficiency of the water heater (namely, how much electricity is required to produce a gallon of hot water). There are only two functions in this particular chain: (1) how much hot water can be produced from a kWh of electricity, and (2) how much utility (i.e. happiness, satisfaction) the individual gets from a gallon of hot water. Since utility is not directly quantifiable, we measure it by how much money the individual is willing to give up to get that increase in utility.

In the case of derived demand for hazardous air pollutant regulation, there are rather more functions involved in the process of characterizing how a given environmental regulation concerning hazardous air pollutants will ultimately affect individual utilities. (The final step, monetization of regulatory benefits, requires the conversion of a given improvement in individual utility levels into an equivalent income difference. This is the topic of a future workshop, so we will leave this final benefit in terms of utility.)

First, we need to outline some of the constituent functions and identify the argument of each function that is of key interest. Each of these functions will of course be multivariate and subject to uncertainty.

(1)

Emj = Em(Ij,…)	Emissions Em from firm j depend upon the firm's inputs and technology, Ij
Ak = Ak(Emj,…)	Ambient concentrations in region k depend upon emissions of all contributing firms j
Exi = Exi(Ak,…)	Exposure of individual i depends on ambient concentrations in their region k
Di = Di(Exi,…)	Dose received by individual i depends on exposure
Ci = Ci(Di,…)	Cases of health effects for individual i depends upon the individual's dose
Si = Si(Ci,…)	Symptoms of individual i depend upon whether they are a victim of health effects
Ui = Ui(Si,…)	Utility levels are probably most directly influenced by symptoms (compromised function, life expectancy, etc.)

Conceivably, each of these relationships could be studied independently. Since controlled experimental data are rare (and may not reflect the true empirical derivatives in the

field), a considerable amount of modeling will be necessary. By formally modeling all of the factors that determining a particular outcome variable, we take other covariates into account and minimize "omitted variables bias" in the estimated slopes. Specifically:

(2)

Emj = Em(Ij,…) - Emissions of a particular pollutant by a particular firm will depend on the firm's inputs, production technology and abatement efforts, including MACT. It may depend upon factor prices (including the prices of precursors of polluting emissions) and upon the prices of the firm's outputs.

Ak = Ak(Emj,…) - Ambient concentrations in a particular region will depend upon the emissions of all firms that contribute to these ambient levels of pollution and upon the "fate and transport" (transfer coefficients) for each firm. Transfer coefficients depend upon weather conditions, season, prevailing winds, the nature of the pollutant, and other factors.

Exi = Exi(Ak,…) - Exposure to the pollutant of individual i will depend upon the individual's behavior and patterns of activity, including avoidance behaviors.

Di = Di(Exi,…) – Dose actually received will depend upon exposure and other factors.

Ci = Ci(Di,…) - Cases as a function of dose level will depend upon the individual's socioeconomic status, current health status, age, gender, and other factors (such as the duration of exposure or cumulative exposure).

Si = Si(Ci,…) - Symptoms, given that the individual develops a case of the health effect, will depend on the individual's metabolism, access to treatment, age, current health status.

Ui = Ui(Si,…) - Utility may depend only upon the spectrum of symptoms the individual does or does not experience. However, it is possible that utility will be affected even if this particular individual is completely asymptomatic. Perhaps knowledge of exposure creates fear, or the exposure and incidence of cases for other people affects utility. There can be both "use" and "nonuse" demands for relief or prevention of symptoms (including fatal cancers).

Studied in isolation, each of these functions presumably has some explicit approximate mathematical form, the parameters of which must be determined from empirical studies. It is not too far off the mark to suggest that each function in the list above is the province of a distinctly different discipline.

Each of the above functions can be embedded into the next. People do not so much want environmental regulation because they want firms to meet MACT standards. Rather, they want environmental regulation because of what it means for themselves and their families. Thus, in a simple model, utility levels depend directly on symptoms being experienced, and indirectly on each of the contributing factors, all the way back to Ii (which we might interpret as abatement

technology), if the EPA should choose to regulate at that level. Consider one possible characterization of individual utility.

(3) $$Ui = Ui (Si (Ci (Di (Exi (Ak (Emj (Ij))))))) $$

The partial derivatives that we need to know in order to "construct" the effect on utility of a change in abatement technology, Ij appear in the following expression:

			∂U_i	∂S_I	∂Ci	∂_{DI}	∂Ex_i	∂A_k	∂Em_j	
(4)	dU_i	0	&&	&&	&&	&&	&&	&&	&&	dI_j
			∂S_i	∂C_I	∂D_i	∂Ex_I	∂A_k	∂Em_j	∂I_j	

If the EPA regulates individual firm emissions:

			∂U_i	∂S_I	∂C_i	∂D_I	∂Ex_i	∂A_k	
(5)	dU_i	0	&&	&&	&&	&&	&&	&&	∂Em_j
			∂S_i	∂C_I	∂D_i	∂Ex_I	∂A_k	∂Em_j	

If the EPA regulates ambient concentrations:

			∂U_i	∂S_I	∂C_i	∂D_I	∂Ex_i	
(6)	dU_i	0	&&	&&	&&	&&	&&	∂A_k
			∂S_i	∂C_I	∂D_i	∂Ex_I	∂A_k	

In the easiest possible world, each of these partial derivatives would be a nonstochastic scalar. This implies extreme linearity in each respective function. But it is likely that most, if not all, of the constituent functions outlined above are rather nonlinear. It is also possible that the derivative of any one outcome with respect to one cause (say $\partial C_i/\partial D_i$ = the effect of a change in dose of benzene on the incidence of cancer) depends on the doses of other HAPs. There can be interactions between stressors that influence the effect of a change in the level of any one stressor.

The benzene, perchloroethylene, and manganese case studies focus primarily on the $\partial C_i/\partial D_i$ link in this very long chain of partial derivatives. But if we are trying to assess the

welfare effects of regulation-induced changes in Ij (or Emj or Ak), these effects depend upon all of the intervening partial derivatives, not just one of them.

But the utility function proposed above is just one of a number of possibilities. It assumes that individuals derive utility from environmental regulation ONLY insofar as it affects the symptoms they experience. It is an economist's job, however, to try to discern just what it is that contributes to individuals' utility levels. It is possible that an individual's utility level depends on emissions Emi not only indirectly via the symptoms they experience from the health effects these emissions create, but also directly on emissions levels. Perhaps the utility function looks more like this:

(7) $$Ui = Ui (Si (Ci (Di (Exi (Ak (Emj (Ij)))))), Ci , Di , Exi , Ak , Emj , Ij)$$

If this is the way an individual's utility level is determined, then the individual may derive an increase in utility from a decrease in emissions or ambient concentrations even if this change in ambient concentrations produces absolutely no health effects ($\partial C_i/\partial D_i = 0$)! Individuals are allowed to derive utility from whatever they want. There is no justification for considering only that utility derived directly from health symptoms.

b) Economists prefer to attempt to value the things that enter most directly into people's utility functions. In the environmental regulation context, this usually means symptoms, such as "days of eye irritation," or "days of moderate cough," or even "statistical lives lost." In the chain of partial derivatives outlined above, we would prefer to explore how people are willing to trade off dollars for changes in the level of symptoms. We might ask them directly what they would be willing to pay, or we might ask them to choose among policies that involve different costs and different levels of symptoms. The dollar metric is merely an intermediate device to capture how much of other things they would be willing to give up in order to achieve a reduction in some set of symptoms.

Focusing on the value of symptom changes reduces the dimensionality of the problem in many cases. This is analogous to the way market researchers sometimes reduce the vast number of different automobiles on the market to a much smaller number of attributes. Each auto can be characterized not by its make and model and year, but by the bundle of attributes that it represents (curb weight, acceleration, MPG, age, number of seats, etc.). The advantage of this method is that if we study the market prices of autos as a function of the bundle of attributes each represents, we can infer how market price depends on attributes. Then, if confronted by a new make and model, with a specified set of attributes (preferably within the range of attributes observed in the estimating sample) we can figure out approximately what people would be willing to pay for the new vehicle.

To reduce the 188 HAPs to a smaller set of spanning "symptoms," these symptoms need to be defined rather grossly, of course. Rather than trying to come up with distinct benefits estimates for changes in the level of each of the 188 distinct HAPs, we would instead endeavor to infer the incremental value of changes in each of a smaller set of symptoms. A one-unit

change in the concentration of a particular HAP would then need to be quantified in terms of what that means for what matters to people: the suite of symptoms they are experiencing (e.g. eye irritation, sore throat, cough, fatigue, for some compounds, and more serious endpoints for other compounds). Policy changes with respect to HAPs that result in a bundle of symptom changes will make it necessary to ascertain whether the effects of distinct symptoms on utility are additive, subadditive, or superadditive.

We would like to stock the research "shelf" with a set of estimates concerning people's willingness to pay to avoid increments in each of the set of symptoms that are the usual suspects in HAP assessments. We would also like to be able to say something about interactions among sets of symptoms. If we can build up this inventory, then we have some hope of reconstructing willingness to pay for a reduction in the amount of a particular HAP or set of HAPs that may not have been studied explicitly. This process is known as "benefits transfer." We do not imagine that it would be possible or sensible to conduct a separate economic analysis of willingness to pay to reduce each one of the 188 HAPs. Instead, we would like to study only a few of them explicitly (and bring in results on willingness to pay for symptom reduction from benefits assessments for criterion pollutants or other applications).

Economists have become more and more confident over the years about how and when they can come up with reasonable point and interval estimates for reductions in symptoms. But we must rely entirely on other disciplines to convert a proposed environmental policy into changes in a set of symptoms for some segment of society. What can we do if other disciplines are unable to specify point and interval estimates for the rest of the partial derivatives? This is only a dead end if we restrict utility to be derived only from indicators of health status.

But keep in mind that even if the most likely magnitude of the dose-response function derivative is "zero" at current ambient concentrations, this does not mean that individuals cannot experience a direct increase in utility simply from knowing, for example, that the ambient concentration of a suspected HAP has been reduced.

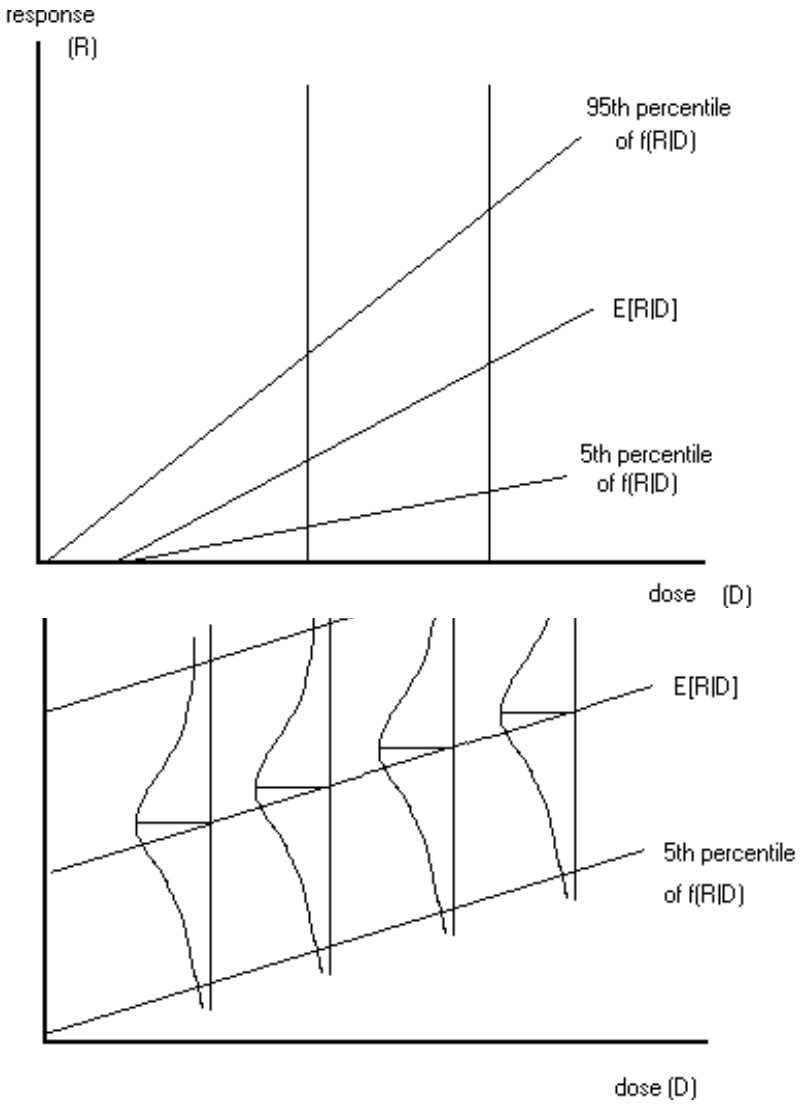
Unlike toxicologists, who rightly expect to be able to identify a mechanism that explains how an increase in the dose of some toxicant contributes to changes in an individual's health status, economists expressly do not seek to figure out how a change in symptoms leads to an effect on individual utility levels. "There is no accounting for tastes." We do not care about WHY people derive utility from something, only that they do. The variant of the utility function that emphasizes only symptoms is consistent with someone having a value system where actions are based on "hypothetical imperatives" in the sense of Kant (i.e. IF you want to achieve this end, then you must take this action). People who care directly about emissions or ambient levels might be viewing HAP control as more of a "categorical imperative" in Kant's terminology (i.e. you must take this action). Individuals can have any of a wide variety of philosophies or value systems driving their individual utility functions. Economists just take these as they are and concentrate on the task of how to aggregate these into a measure of collective social welfare that respects these individual utilities.

c) Why do economists want to know about central tendencies? Can't they figure out what they know from the information about the 95th percentiles???

Since my logistic dose-response curves would be so untidy as to obscure the point, I will draw linear dose-response relationships.

If the conditional distribution of responses at a particular dose had the same shape as the conditional distribution at any other dose (only with a different expected value), then the change in the conditional expectation of the response distribution for a given change in the dose would be identical to the change in the 95th percentile of the response distribution for the same change in the dose. It would not be necessary to know the central tendency.

Are the conditional distributions of responses identical at all dose levels? This is an empirical question. It may be convenient to assume that they are, so that the needed slope (derivative) is not only constant, but the same at all percentiles of the conditional distribution. However, without evidence to support this rather heroic assumption, it is more reasonable to allow for a conditional distribution of responses that varies across dose levels. Without drawing the precise shapes of each of these conditional distributions, the lines that connect the 95th percentiles, the expected values, and the 5th percentiles could just as easily look like the second diagram. Here (below), the assumption of linearity in the percentiles is retained, but even this may be untenable.



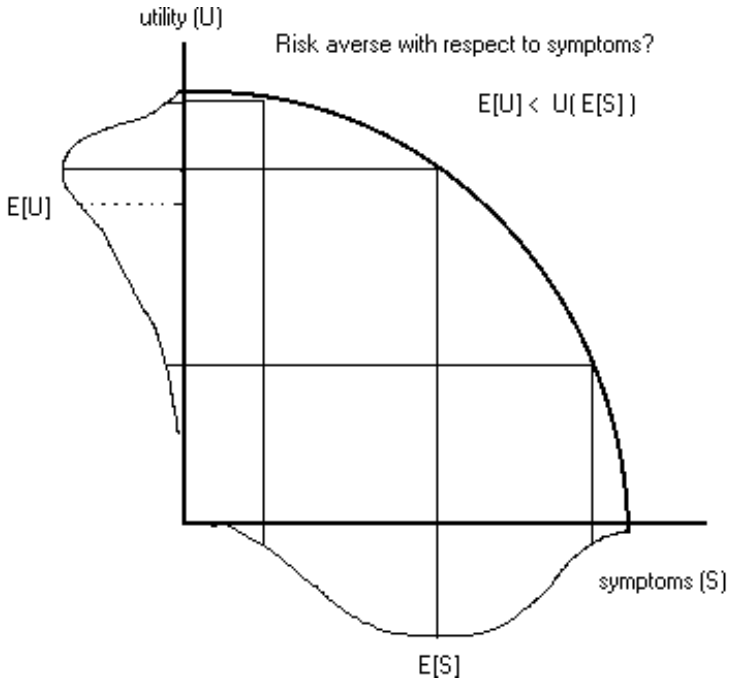
In this more general case, it is clear that the derivative implied by the relationship between expected response and dose will be quite different from the derivative implied by the relationship between the 95th percentile and dose.

Knowledge of the profile of the 95th percentiles of the conditional distribution of response, given dose, is insufficient to determine the complete distribution. Even a normal distribution requires two moments to identify its exact shape. Usually, these are the mean and the variance. But you could equally well pin down a specific normal distribution with information about its 95th and 5th percentiles (because the distribution is symmetric, this implies the location of the mean, and the range between these percentiles implies the dispersion. The 95th percentile alone is insufficient to identify the distribution, even if it was known to be normal (or lognormal.) To be able to identify a distribution based on a single percentile, we need a one-parameter distribution. An example would be the exponential distribution. The location and scale of an exponential distribution would be completely conveyed by the information in its 95th percentile. But can we assert that the conditional distribution of responses, given dose, is exponential. Not without evidence.

d) Why do economists need conditional distributions of response, given dose? If we are going to do a rigorous assessment of the overall effects on benefits from environmental regulation, we need to know both the central tendency and the dispersion in each of the partial derivatives that must be multiplied to produce an estimate of utility gains from HAP regulation. Since the product of random variables is unlikely to have an analytically tractable distribution, simulation methods will typically be required to generate a distribution for benefits that allows the reporting of not only best estimates, but also comprehensive error bars.

e) Utility levels under uncertainty. The problem with a utility function that is linear in an uncertain symptom is that it does not allow for risk aversion with respect to that symptom. A deviation from the expected symptoms that results in lesser symptoms may not have nearly the

negative effect on utility of a deviation in from expected symptoms that results in worse symptoms. The loss of utility when the outcome differs from its expected value may be very much asymmetric on either side of the expected value of the outcome. There has not yet been much empirical work on alternative kinds of utility function for use under uncertainty. Further research on how to implement benefit-cost analysis under uncertainty in more general contexts is clearly needed. Relevant considerations include how individual subjective uncertainty is distinct from scientific uncertainty or disagreement with respect to the ultimate symptom differences to be expected from a HAP regulation.



The associated figure shows a utility function with respect to symptoms, U(S), for an individual who is risk-averse with respect to symptoms. We need to know the distribution of symptoms in order to figure out the distribution of possible utilities associated with these symptoms. Note that a utility function characterized by risk aversion means that a symmetric distribution for S is converted into a skewed distribution for U. The important result is that expected utility E[U] is not simply equal to utility at the expected level of symptoms. It is likely to be less. How much utility is lost due to uncertainty about the level of symptoms depends on the shape of the distribution of S and on the shape of the utility function.

Economists have a pretty good theoretical framework for Benefit-Cost Analysis under uncertainty. A succinct exposition of the theory in the case of objective uncertainty over a binary outcome was presented by Graham (1981). For simple utility functions (e.g. those which are linear in the levels of any factor that is uncertain) it is a straightforward matter to generalize that model to the case of a continuum of possible outcomes (such as the uncertain response to a dose of a HAP, or more directly, the uncertain symptoms from a dose of HAP). When utility is depends upon the level of a symptom AND upon the squared deviation of the symptom from its most likely value, the mathematics is simple and expected utility depends on the expected level of the symptoms and on the variance in symptoms around this expected level.

f) But we do not live in an ideal world. Unfortunately, the perfect ingredients for a full benefit-cost analysis under uncertainty are not available. Some of the key derivatives in the chain are unknown and probably unknowable at finite research cost. How, then, can an economist begin to measure the benefits that people derive from regulation of hazardous air pollutants? Rather than the "bottom-up" approach of building the benefits from derived demand for regulation via demand for reduced health risks, we can consider a more "top-down"approach. This approach accommodates both limited scientific information and individual consumers' subjective assessments of the key partial derivatives (or even just the products of these partial derivatives). I make the following suggestions for a research strategy based on what I am currently attempting to do in the case of derived demand for climate change mitigation policies Cameron (1998). Understanding and measuring the demand for expensive climate change mitigation programs has much in common with the problem of understanding and measuring the demand for HAP regulations. Both issues are characterized by uncertainty (incomplete or ambiguous scientific information, controversy among experts, competing corporate and advocacy group positions on the science, varying levels of expert "credibility," and varying degrees of public interest cross-sectionally) and long latency periods.

The individual's subjective distributions on the magnitudes of the partial derivatives in the chain are a product of the interaction between their experiential knowledge and the information that they receive from outside sources. For example, if the individual knows somebody who died of cancer who worked in a dry-cleaning establishment, this knowledge will be combined with whatever expert information to which the individual has been exposed to yield that individual's *subjective* conditional distributions of responses in the dose-response relationship between perc and cancer.

Thus, the proper characterization of individuals' willingness to pay for HAP regulation depends on their subjective assessments of all of the relevant partial derivatives (either individually or in compounded form). An appropriate research strategy would elicit from each individual their *subjective assessments* of what HAP regulation would be likely to achieve in terms of health effects (and other effects). If he or she desires it, the individual should have access to summaries of whatever expert information is available, including the fact that this information is complete, if that is the case. The individual's updated *subjective assessment* should then be established, and then his or her preferred choices among a set of policy alternatives should be elicited. These policy alternatives should differ not only in terms of their costs to the individual, but also in terms of the degree of protection provided against the consequences the respondent expects without regulation. This method accommodates consumer utility from HAP regulation no matter how it arises, whether through avoidance of measurable health effects or simply through "existence" demand for a non-toxic environment. Economists do not presume to question where utility enhancements come from, only whether they exist.

This method of assessing the benefits of HAP regulation is likely to rely upon stated preference techniques (i.e. contingent valuation or its generalizations), since it is hard to imagine actual referenda being held on alternative HAP policies. Fortunately, researchers are

understanding more and more about the limitations and idiosyncracies of stated preference research and how to minimize these.

It is important to keep in mind that economic research concerning the apparent value of HAP regulations conditional on the public's understanding of the risks these compounds present *does not mean that we have to make policy based on widely held misperceptions about the true risks of HAP concentrations*. The idea is to model benefits as an explicit function of *perceived* risks. Once this function is understood, it is then straightforward to replace the *subjective* risks with *scientifically supportable* levels of objective risk and re-calculate the implied level of benefits that would accrue to each individual if their beliefs were consistent with the science. Note that the science can still involve uncertainty, and the values that individuals are "predicted" to hold for HAP regulation should definitely be conditional on the extent of uncertainty (either individual subjective uncertainty, or scientific uncertainty).

g) A smattering of philosophy. Benefit-Cost Analysis, as it is interpreted in most contexts today, is understood to be based upon a utilitarian (Benthamite) social welfare function. It is important to keep in mind that this particular social welfare function is not the only game in town, although most economists in the U.S. are steeped in the utilitarian tradition because it makes the benefit-cost problem tractable and it does have a number of desirable properties. (See Kolstad (2000).)

Imposing an environmental regulation would be a "no-brainer" if it made nobody worse off and at least somebody better off. Then nobody would be opposed to it and at least one person would be in favor. Environmental regulations are controversial only if the beneficiaries gain at someone else's expense. Some individual utilities will go up (or the regulation would not be demanded) but other people's utilities will go down. This is true, for example, if the beneficiaries of the regulation are not the same people that bear the costs.

Suppose there are some winners from regulation and some losers. If the winners win big enough to be able to compensate the losers for their losses, then it would be possible to achieve unanimity about the desirability of a regulation.

Just before WWII, Nicholas Kaldor and John Hicks proposed that the secondary consideration of the distributional consequences of some proposed reallocation of resources can be separated from the primary discussion of whether the net change in utility is positive overall. Even if compensation does NOT take place, if the "benefits" in terms of gained utility for some members of society exceed the "costs" in terms of lost utility for other members of society, then there is an argument that the proposed reallocation is a good idea for society as a whole. It is a second-order issue to then consider whether the distributional consequences of the reallocation are sufficiently undesirable to preclude the reallocation on distributional grounds. This so-called "compensation principle" says that if a proposed reallocation would create more gains than losses, then it is socially desirable, even if no compensation occurs.

A Social Welfare Function (SWF) takes the utility levels of individuals and combines them in some fashion to yield a single-dimensioned scalar number for something called "social welfare." Different value systems lead to different candidates for the SWF.

Utilitarian (Benthamite): $W(u1,\ldots,uN) = \sum_i s_i u_i, \qquad s_i \geq 0$

Aggregate social welfare is a weighted sum of the utility of each individual in society. The weights are positive, but need not be equal. We are usually most interested in welfare *changes* from *resource reallocations*: $\Delta W/\Delta x$ (where x is a determinant of individual utilities, ui). This is a way of denoting the "net benefits" from resource reallocation "Δx." If these net benefits are positive, the utilitarian (Kaldor-Hicks compensation principle) opinion would be that the reallocation is a good idea. This is Benefit-Cost analysis, as conventionally practiced.

Egalitarian: $W(u1,...,uN) = \sum_i u_i - \theta \sum_i [u_i - \min_i(u_i)] \qquad \theta > 0.$

In this SWF, society cares about the total amount of utility, $\sum_i u_i$, but also about the degree of inequality. If everybody enjoyed identical utility, the term $\sum_i [u_i \min_i(u_i)]$ would be zero. The negative sign indicates that social welfare is decreased by departures between individual utilities and the lowest individual utility level. The θ parameter dictates the weight on distributional issues (i.e. inequality).

Rawlsian: $W(u1,...,uN) = \min_i(u_i)$

A society is only as well off as its least fortunate member.

i) A reminder: Arrow's Impossibility Theorem (AIT). This was one of the discoveries for which Kenneth Arrow won the Nobel Memorial Prize in Economics in 1972 (shared with Sir John Hicks of Kaldor-Hicks fame). The result was part of his Ph.D. dissertation. To paraphrase its useful result: *There is no "ideal" way to combine individual preferences into a social choice mechanism*. Since then, people have spent a lot of time tweaking the conditions, trying to figure out under just what modified conditions you CAN produce a nice tidy theory of social decision-making.

But this impasse cannot stop the frequent need to make policy decisions regarding the allocation of resources in some "best" fashion. In practice, the Kaldor-Hicks compensation principle is often used, with ex post consideration of the severity of the distributional consequences. The AIT just reminds us that this decision is not necessarily the only, or necessarily the best, way of deciding about resource reallocations.

Some Criticisms of the Utilitarian Approach to Decisionmaking:

1) Do we each have a utility function that adequately and consistently represents our preferences, especially over time. Tastes change. Preferences can be manipulated by information campaigns (e.g. advertising, public service announcements, "education").

2). Decisions according to the benefits and costs experienced by current members of society are suspect if not all of the affected individuals are taken into consideration. Specifically, future generations are sometimes not represented. Their tastes may differ from current generations.

Should public policy be based on individual preferences (consumer sovereignty) at all, or on what is morally right? Utilitarianism is a branch of teleological ethics. An alternative ethical system is deontological ethics (deontology= science of duty; moral obligations). Immanuel Kant (*Metaphysics of Morals* (1785) considered that actions can be judged by their intrinsic "rightness" and not by the extent to which they serve to further one's goals or aspirations. Two types of imperatives direct our proper behavior:

1) *hypothetical imperatives* (present the practical necessity of a possible action as a means of achieving something else which one desires (or which one may possibly desire; e.g. IF you want that, you must do this!)
2) *categorical imperatives*: present an action as of itself objectively necessary, without regard to any other end. e.g. You must do this!

If you subscribe to a deontological ethics, you would not require a benefit cost analysis to justify environmental regulation, you could justify it solely on the basis of an argument that "humans have no business messing up the environment with HAPs." The major problem in using deontological ethics as a basis for policy is that reasonable people can differ in terms of what they judge to be "intrinsically right." These systems can work pretty well in a homogeneous society, but the more heterogeneous the society, the more difficult it is to agree on what constitutes an intrinsically right course of action with respect to policy. We rely on utilitarianism because it makes policy evaluations easier to effectuate. (See Hackett (1998).)

(Some, and perhaps much, of the animosity towards economic welfare analysis stems from misunderstandings about what it IS. There is a vitally important distinction between using Benefit-Cost Analysis to MAKE environmental decisions, versus using Benefit-Cost Analysis to INFORM environmental decisions.)

j) Summary. As an economist who has struggled for quite some time with a number of different problems in valuation of non-market benefits of environmental goods, I really appreciate having had an opportunity to get confirmation from toxicology experts that there are fundamental (and probably unresolvable) gaps in our knowledge about the measurable health consequences of hazardous air pollutants.

This insight means that an unambiguous, *objectively calculated*, bottom-up measure of the social benefits from HAP regulation is unlikely to be forthcoming. But this certainly does not mean that these benefits are zero. People may be willing to allocate society's resources (including their own) to control of hazardous air pollutants simply because there is a possibility that they could have health effects, even if these have not yet been detected (or are unlikely to be detected

any any feasible cost of scientific research). Deciding upon an alternative valuation strategy is the next step.

It seems likely that stated preference research will be the most fruitful way to proceed. Experts in economics and cognitive psychology will have to address the issue of elicitation of subjective probability distributions for health risks (and other risks) that may emanate from ambient levels of hazardous air pollutants. We will need to study how public perceptions of risk can be modified by new scientific results, or by "education" (propaganda) campaigns. People's values for environmental programs depend upon what *they think* the programs are buying them. Their beliefs may or may not be consistent with current scientific understanding. But this does not preclude a strategy of first elicitation, and then simulation, to ascertain what *would have been* the public's value of a specified program had they fully accepted the best current (and possibly incomplete) scientific knowledge.

References

Cameron, T. A. "Credibility of Information Sources and the Formation of Individuals' Option Prices for Climate Change Mitigation." . Department of Economics, UCLA, June 1998.

Graham, D. A. "Cost-Benefit Analysis under Uncertainty." American Economic Review 71, (1981): 715-725.

Hackett, S. C. Environmental and Natural Resources Economics. Armonk, New York: M.E. Sharpe, 1998.

Kolstad, C. D. Environmental Economics. New York: Oxford University Press, 2000.

G.4. Comments of Ms. Laurie Chestnut, Stratus Consulting, Boulder CO

a) Comments relevent to workshop agenda questions 1 and 2:

Question 1) Proposed approaches for hazard assessments for selected HAPs that would facilitate benefit assessments for those chemicals.
Question 2) Expert discussants' views on whether it is possible to produce a methodology for developing central tendencies and distributions in hazard assessment for HAPs for use in benefits analyses and how that might best be done\

It seems pretty clear from the information presented at the workshop that available toxicology and epidemiology evidence is insufficient to provide dose-response relationships over relevant ranges of exposures for most of the air toxics for which EPA is required to make regulatory decisions. Data based on human studies tend to be at much higher occupational exposures, requiring heroic assumptions when extrapolating to typical community exposures. Data based on animal studies are suggestive of whether or not, and by what mechanism, a chemical may be harmful, but are very difficult to interpret when it comes to quantitative dose

response for humans. Al McGartland gave references to some published papers that look at drawing dose response from information used to determine reference doses for some chemicals. These should be reviewed by the toxicologists to see if the methods are sound and how widely they could be applied. It appears, however, that the information necessary to do a comprehensive quantitative benefits assessment of air toxics reductions, such as is the goal for the 812 assessment, is not going to be available any time soon.

Some questions arose at the workshop regarding if and how economics would use information on variations in dose response for different population groups. Heterogeneity in dose response is a part of the dose response information needed for a benefits assessment. For each dose response function, we need to know to what population it applies. It does not need to be for the general population. There may be different dose response for different age groups, for example. If certain groups are affected significantly differently than others, this should be included in the presentation of the results of the assessment. Similarly, it is important to know whether, for example, there are 100 people facing a change in risk of cancer of 1 in 100 or a million people facing a change in risk of cancer of 1 in a million. The bottom line in both cases may be one expected cancer case, but the risk situation is quite different.

b) Comments relevent to workshop agenda questions 3 and 4

Question 3) How best to identify limitations and uncertainties in both risk assessment methods and economic models.

Question 4) Suggestions and prioroities for a research agenda to address identified gaps in available data and methods needed to conduct HAPs-related benefit analyses

In the regulatory decision making context, there is strong motivation to make reasonable use of the available information, even if all the important questions cannot be answered. In this context, ranges of estimates of risk changes, upper or lower bounds on risks, and other information short of a best estimate of dose response, can be utilized to help inform regulatory decisions. Highly variable, uncertain, and inconsistent information, however, is difficult to use in a quantitative benefits assessment unless there is some way to assess how likely it is that each result is accurate. For example, if some studies have found that a given chemical is a carcinogen, and others obtain negative results for the same chemical, we can say that the expected reduction in cancer cases ranges from zero to the amount suggested by the studies that have found an association. Such inconsistencies in results can result in such a large range in benefits estimates that the assessment is not very useful (e.g., saying that the benefits are somewhere between zero and $10 billion is not usually very useful information for decision makers). However, if some assessment of the likelihood that each of the available results is accurate can be made, then the assessment can be made more informative. Continuing with the same example, perhaps there may be a basis for determining that studies finding a carcinogenic effect are more likely to be correct than studies that have not found carcinogenic effect. Thus, it may be possible to say that there is a 75% chance that the benefits are $10 billion and a 25% chance that the benefits are zero. This

does not mean that the only benefits number that should be presented is the $7.5 billion expected value. The range and the probabilities should be presented.

Probability distributions (or some other form of assessment of the likelihood that various results are correct) are also very useful when combining the many steps involved in a benefits assessment. When many ranges of values are all multiplied together, the high and low values can end up being very far apart. A very wide range in the result is not very useful information, especially when the chances that all the low values or all the high values are correct is small. For example, if we take the highest value at each step and multiply them all together, we get a very high value result that is very unlikely to be correct. The ideal information needed to determine probability distributions on results used as inputs for benefits assessment is seldom available, and some professional judgment is inevitable. Sometimes it is better to say we don't know than to rely on pure guesswork. Where that line is is also a matter of judgment. Sometimes the best we can do is some simple sensitivity analyses on key assumptions in the assessment. For example, if this chemical is not a carcinogen, the answer is X; if it is a carcinogen the studies showing an effect suggest the answer is Y.

G.5. Comments of Dr. A. Myrick Freeman, Department of Economics, Bowdoin College, Brunswick, ME

The question posed for us was "whether it is possible to produce a methodology for developing central tendencies and distributions in hazard assessments for HAPs for use in benefits analyses..." I think that in short the answer is "No," because of both the variety of endpoints of concern (cancer and a variety of non-cancer endpoints) and the variety of sources of data (human epidemiology, long term animal feeding studies for cancer, and other human and animal data for those non-cancer endpoints that have been studied). Rather, different approaches will be necessary for different endpoints and types of data.

For the 20 or so known human carcinogens, best estimates of dose-response (D-R) functions and uncertainty bounds can be obtained using meta-analyses or applying Monte Carlo methods to the available human data. This leaves us with the questions of extrapolation to low doses and the possible existence of thresholds; but these are well known questions.

For the other possible carcinogens, one possible answer is to try to obtain better human data on D-R relationships; but there are the well known problems of estimating human exposures and the typical low power of epidemiology studies. Lacking additional human data, the only recourse is to the animal data. This would mean obtaining maximum likelihood estimates of cancer slope factors and their confidence intervals from the original animal test data. Again there is the question of high to low dose extrapolation as well as the animal to human concordance and extrapolation questions. But I don't see any other way to proceed. I second Lester Lave's suggest concerning the comparison of animal and human data for those substances for which both kinds of data exist.

Resorting to the animal data for all of the possible HAPs carcinogens that lack human data is probably impractical. It will be important to perform some kind of screening and prioritizing exercise based on, e.g.,, indications of large volumes of emissions or where human exposures are thought to exceed RfCs. Screening the noncancer HAPs can proceed in a similar fashion. For example, where present human exposure is less than the RfC, the benefits of further reductions in emissions are likely to be zero.

The preceding advice has been based on the assumption that benefits are being defined in the standard way as willingness to pay (WTP) for reductions in the numbers of cases of various types of disease, eg., cancer, obstructive lung disease, etc. An alternative approach to benefit estimation (which might be pursued in parallel with the standard approach) is to investigate different ways to define the commodity to be valued that reflect the ways in which individuals actually think about reducing the risks of environmental disease. For example, economists have used the averting behavior model to analyze data on bottled water purchases as a way of valuing reductions in the risk of waterborne disease. My conjecture is that many individuals view their purchases of bottled water, not as a means of reducing the risk of specific diseases, but as a means of increasing safety more broadly conceived. If there is anything to this conjecture, then individuals might view a broad or comprehensive policy of controlling emissions of HAPs as producing safety or "peace of mind" rather than as yielding reductions in the risks of specific diseases. And if that is the case, then best estimates of reductions in specific risks are not required for benefit estimation. What would be required, however, is a better understanding of the relationship between controls on emissions of HAPs and individuals' perceptions of safety or "peace of mind."

G.6. Comments of Dr. Dennis Paustenbach, Exponent, Menlo Park, CA

Response to workshop agenda Question :2: expert discussants' views on whether it is possible to produce a methodology for developing central tendencies and distributions in hazard assessments for HAPs for use in benefits analyses and how that might best be done.

The answer to this question depends on the level of certaintythat one needs to satsify those who will perform and use benefit/cost analyses. Certainly, we can produce cancer risk estimates for animal and human carcinogens using various models. We have used these models in the past in an attempt to rank (in a relative way) the carcinogenic potency of chemicals. However, like most risk assessors, I don't believe that they can accurately predict the actual human response following low exposure to these substances.

I acknowledge that a couple published analyses have suggested that for the genotoxic chemicals, where we have exposure and epidemiology data, there has been a reasonable level of agreement between the number of cases predicted vs the number observed. However, the doses in these studies were those observed in workplaces of nearly 30-40 years ago and these are not even within 100 fold of the doses to which the public is exposed today. In addition, for the cancer estimates from the basic models to be accurate, they should be attached to the "internal" doses estimated by physiologically-based pharmacokinetic (PB-PK) models. This has been best

exemplied in a paper by Reitz et al (l996) which studied vinyl chloride. In short, for virtually all of the animal carcinogens, we are not in a position to use low-dose models to predict the actual cancer risk in humans. We can, for purposes of benefit/cost analyses, use them as a relative index of hazard which could then be "modified" if one wanted to consider other biological factors (like genotoxicity).

Thus, the answer to the question "is it possible to produce a methodology for developing central tendencies and distributions" is yes for the carcinogens. This could be done and, given the abovementioned caveats, it might serve the purposes of the exercise.

With respect to the non-carcinogens, a different approach would be needed as there is assumed to be no risk associated with doses below certain values. One possible approach is to calculate "margins of safety" (MOS) for the non-carcinogens. In this approach, one would take the EPA Reference Dose (RfD) or the Reference Concentration (RfC) and determine the cost associated with achieving doses below these "safe" concentrations. As was mentioned by Dr. Lave, often the public simply wants "to feel safe'....no matter that scientists may give assurances that current conditions pose no significant risk. Assume, for example, background concentrations of formaldehyde in some cities can reach, under certain conditions, about 50 ppt. An airborne concentration generally thought to pose no risk of even transient eye irritation is about 250 ppt. Perhaps, the public wants then never to have concentrations get above 25 ppt. The economists could then provide information to Congress which indicates that the cost of providing a MOS of 10 for formaldehyde is $100,000,000. A decisionmaker can then compare this cost to achieve an MOS of 10 for formaldehyde to the cost of achieving an MOS of 10 for another non-carcinogenic chemical, for example manganese, and could then weigh the relative importance of the potential adverse effects. In the case of formaldehye, the threat is transient eye irritation, while for manganese, it could be premature aging of the nervous system. If the costs were similar to achieve an equivalent MOS, then it is likely that the decisionmaker would choose to regulate manganese to rather than formaldehyde given the major differences in toxicity. This approach doesn't equate to a "life save" but should be a perfectly useful metric for both the economist and regulator.

As shown in the above example, it is necessary that the process of benefit-cost analyses for the HAPs be tackled according to the adverse effect of concern. For example, probably about 33% of the chemicals are listed due to their carcinogenicity, 33% are listed because they are systemic (non-carcinogenic) toxicants, while about 33% are irritants. Each might require a slightly different approach. Nonetheless, each has a dose-response curve (or one could be built) and a distribution around the various points on the dose-response curve could be built (for both the carcinogens and non-carcinogens). Over time, the process would almost certainly be modified as more is learned about its strengths and weaknesses.

G.7. Identifying Limitations and Uncertainties in Risk Assessment and Benefit Measurement Methods. Comments of Dr. V. Kerry Smith, Center for Environmental and Resource Economics Policy,Department of Agricultural and Resource Economics, North Carolina State University, Raleigh, NC

This paper is intended as a summary of some of the issues raised in the SAB/EPA Workshop on Benefit Analysis for Policies that reduce exposures to Hazardous Air Pollution (HAP). My specific focus is on the research needed to address limitations and uncertainties in risk assessment (RA) and benefit measurement (BM) methods. The paper is developed in five short sections after this introduction. The first describes a few features of EPA's conventional practices linking RA and BM. The second explains how the issues posed by HAP are different. In the third I define consequential uncertainty and how this concept may offer an approach to help conceptualizing some of the research needed in this area. Section four discusses some activities in risk assessment and benefit measurement where there appear to be opportunities to co-ordinate research. The last discusses the merits of a parallel research strategy.

a) Conventional Practice. The logic linking RA and BM in the evaluation of criteria air pollutants has been unidirectional. For each air pollutant RA estimates a set of health outcomes such as changes in the probability of premature deaths for the general population, or for specific sub-groups (e.g., elderly, asthmatics, etc.), with a well-defined change in each group's exposure to a specific air pollutant (see Chapter 5 and Appendix D of U.S. Environmental from reductions in the ambient concentrations of a pollutant rely on developing benefit Protection Agency [1997] as examples). Monetary values for the health benefits arising measures for unit changes in each health outcome or risk change. For example, if a reduction in the ambient concentration of particulate matter reduces the expected number of days with respiratory illness, then the monetization of the value of this change generally seeks a measure of the benefits of avoiding a day of respiratory illness. These unit benefit measures may not be associated with the specific source of the health effect. As a result, this approach assumes that a day of respiratory illness is essentially the same regardless of what caused it and therefore would have the same economic consequences for affected individuals.

The unidirectional logic is important. It establishes the equivalent of a chain of functions linking emissions to the ultimate health outcome, and, from the economists' perspective, the change in well-being experienced by each individual (see Trudy Cameron's paper for a more detailed elaboration of this logic).

Any description of the limitations and uncertainties in existing practice usually begins by distinguishing uncertainties that arise because many of the components of the chain of relationships linking emissions of pollutants to changes in well-being are stochastic processes. Reductions in the mortality effects of a pollutant, for example, represent changes in the probability of premature death for specific groups of individuals. Because the outcome is a probability change, the framework used in this case to measure the health effect acknowledges that there are many external and internal influences on an individual's probability of dying in a given year. Exposures to specific air pollutants are only one class of these influences.
The specific models used to fill in the logical chain may incorporate the inherent stochastic nature of the process or they may deal with expected values, implicitly assuming the relationship can be treated as exact except for measurement errors. This distinction determines whether the analysis acknowledges what one might term "structural uncertainty".

A second important source of uncertainty will be labeled "estimation uncertainty". This source arises because one must estimate each of the functions comprising the logical chain of linkages from the emission to the change in people's well-being. Measurement error, omitted variables, incorrect functional form, etc. can introduce error and therefore uncertainty in the actual policy description used to implement the chain of linkages from emission to benefit. While in practice we cannot separate these two sources of uncertainty, at a conceptual level it is generally important to distinguish them.

An important lesson that has been derived from two decades of research developing RA and BM methods for air quality regulations is that health outcomes identified in the RA process must be capable of being associated with "things" people can, in other contexts, choose. For example, small changes in mortality risk can be selected by people in a number of types of behavior. The most common source of behavioral information used in benefit measurement is in job choices. Increased risks on the job generally lead to compensating differentials in wage rates (see Viscusi 1993]). Changes in a person's physical condition, e.g., difficulty in breathing, increase in hypertension, do not generally associate with a choice we can observe people making in another context. In these circumstances a new link in the chain must be introduced. That is, the analyst must associate a change in lung functioning with greater incidence of sick days or an increased change of more serious respiratory illness and then evaluate whether there are actual or stated choices that can be related to these outcomes.

This need to augment the chain of linkages is important because to the extent there is discretion in either the intermediate measure of the health characteristic sought or the observable health outcome that results, risk assessment decisions should be made in ways that facilitate making connection to an economic choice. Otherwise, one simply adds to the uncertainty in the measurement process.

Co-ordination between the design of risk assessment and benefit measurement methods is important for another reason as well. Benefit measurement methods rely on observing (or offering potential opportunities) for people to make choices that reveal how they would tradeoff some outcome that can be related to the pollutant of interest and another commodity that can be expressed in monetary terms. The unidirectional logic of most benefit information for air pollution policy assumes that these choices can be observed in another context and transferred to any evaluation of policies involving air pollutants. Expressed in terms of the chain of linkages discussed by Cameron, and illustrated in simple terms in Figure 1, this process allows monetization of the changes in outcomes at step 5 in the process as approximate measures for the value of their consequences for well being identified as step 6 in this logic.

Figure 1:Chain of Linkages for Environmental Policy Evaluations for Air Pollutants

Characterization of:	Step Number:

Emission Rates for Pollutants by Source	1
9	
Spatial Diffusion, Atmospheric Interactions, and Other Influences to Ambient Concentration of Pollutants	2
9	
Exposure Patterns for Receptors at Each Location	3
9	
Physical Response to Exposure to Pollutants	4
9	
Health Outcome Resulting from Physical Response	5
9	
Effect on Individual Well-being	6

Unfortunately the unidirectional flow is itself an approximation and the very behavioral choice relied upon to make benefit measures for health outcomes can affect the reliability of the one way flow of causation. That is, revealed preference arguments assume to the extent people recognize the effects from steps 2 to 6, they will react to them and try to adapt.[1] The hedonic model assumes they consider site specific amenities in residential location choices. Models of averting and mitigation behavior suggest that other, less costly, responses will also be made. These can include spending less time outside during high pollution times or purchasing central air conditioning, etc. These choices imply the equivalent of feedback loops between steps 6 and 3. The physical responses to some types of pollution may be more easily recognized by people. As a result, in this case the mitigating responses may be more likely and the feedback important. This dimension also will be important to the RA/BM connections with hazardous air pollutants.

b) Three case studies at the SAB/EPA Workshop illustrated the range of possibilities in information likely to be available about hazardous air pollutants. Benzene offered a case with substantial information (compared to other HAP's) on the relationship between exposure (at high doses) and human health outcomes. While there was little basis for evaluating the extrapolation to low dose levels, there also was no basis for arguing modifications to conventional practices.

---

[1]These averting or mitigating behaviors imply continuous adjustment is feasible. To the extent the set of available choice alternatives is finite, then, as Bockstael and McConnell (1999) suggest, people will select the best alternative in the set. This does not imply the choice will be at a point where the marginal value of the amenity underlying the choice equals the marginal cost of adjusting to obtain it.

By contrast, perclorethylene illustrated a situation where despite substantial data for humans and animals, the strategies for measuring impacts were incomplete due to data limitations. Thus the evidence available could only be regarded as "weak signals" of potentially more important effects. Nonetheless, over the short term there seemed to be little basis for improving the information available and some perception that judgments connecting exposures to health outcomes would need to rely on encoding experts' judgments rather than more formal empirical tests.

The last case, manganese, had most of its information concentrated in describing aspects of step 5 in the linkage chain given in Figure 1. Completing the linkage required judgments to connect physical responses to more conventional health outcomes and to evaluate how economically meaningful choices could be connected to non-traditional outcomes.

An overall implication of the background presentations by both Farland and Lave was that policymakers are unlikely to have the type or the level of detail in information available for the HAPs to be evaluated. Moreover, decisions about regulating them will be made before the research required to implement the conventional RA/BM logic would be available.
Three issues emerged implicitly or explicitly in the resulting discussion: screening rules across HAPs were needed to identify the most likely candidates for regulation; the methods used to conduct RA and BM need to consider the treatment of uncertainties in the component elements of policy analyses evaluating practices based on they become "consequential " for decisions; and future research programs should be structured to include parallel research activities, creating pathways for cross-checking findings. The first of these is discussed in comments prepared by Chestnut and Locke; the second is discussed in the next section; and the last is considered in Cameron's comments and in the last section of this paper.

c) Consequential Uncertainty. Regulatory policy based on risk assessment by definition recognizes that a policy is intended to change the stochastic environment in which lay people must make their decisions.[2] Thus, a policy evaluation of a regulation in this context describes uncertainties people face with and without a regulation. Estimation and structural uncertainties are both reflected in such descriptions.

The treatment of uncertainty becomes consequential when it alters one or more aspects of the criteria influencing policy decisions in a way that would alter a choice. In short, analysis decisions about how to reflect the sources of uncertainty for HAP don't matter if the policy choices would not be affected. Thus, improvements in Monte Carlo simulation or other analytical details that would improve variance estimates for health effects are not consequential, if policy choices are always based on central tendencies and these conclusions would not change with the refinement.

---

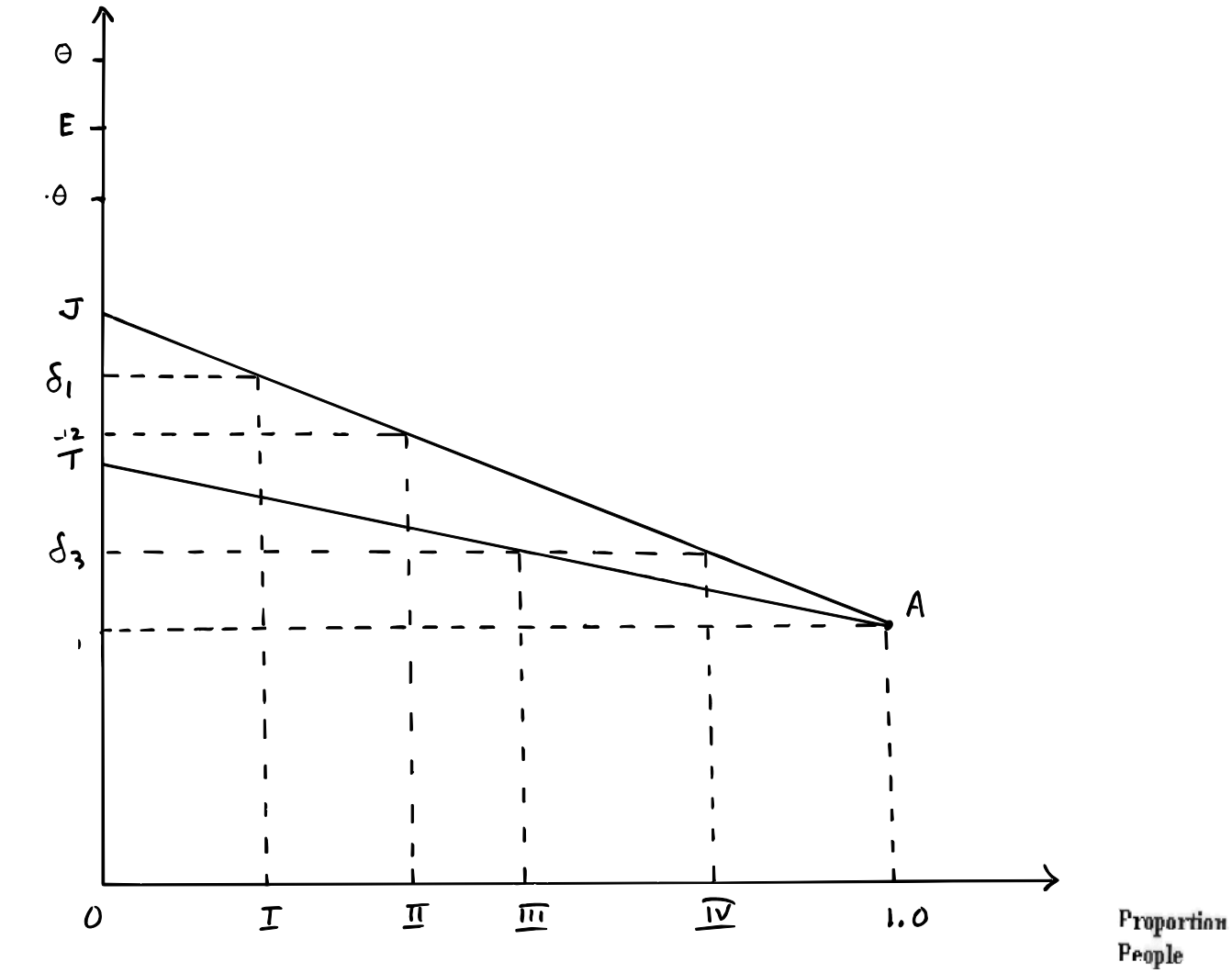
[2]I will not attempt to discuss in this short paper subtle distinctions sometimes considered in economics (and psychology) between the terms risk and uncertainty.

Figure 2 illustrates how the decisions made about coordinating risk assessment and benefit measurement influence whether the treatment of uncertainty is consequential to policy decisions. Five issues contribute to this judgment: the baseline distribution of exposures people receive in the absence of action; the ambient concentration judged to be associated with health outcomes that "count" for regulatory purposes; the estimation uncertainty in describing that ambient concentration; and the risk factor applied based on EPA's propensity to include a margin of safety (due presumably to a composite of concern about structural and estimation uncertainty) complete the elements usually associated with risk assessment. Measures of the economic importance (e.g., unit benefit measures) complete the interacting factors that, together with cost, determine whether regulatory decisions will be consequential to policy choices.

Using Figure 2, on the horizontal axis is plotted the proportion of people experiencing at least the amount of a pollutant measured on the vertical axis.[3] The point A0 designates background, so 100% of the population experiences at least this level. For each HAP we would expect a different relationship to describe these exposure profiles. Different population groups could also be represented as having different patterns of exposure. Curves AT and AJ illustrate two alternatives. They are presented as straight lines for graphical convenience only.

[3]This distribution could refer to the whole population or a specialized group such as children or the elderly.

Figure 2: Illustration of Effect of Treatment of Components of RA and BM for "Consequential" Uncertainty



Policy choices to recognize a health outcome (whether traditional or non-traditional) and link it to an ambient concentration implicitly solve "backwards" or invert the functions one associates with the connections implied by the ambient concentration Y exposure level; the exposure level Y physical response; and physical response Y health outcome linkages (steps 2,

3, 4 and 5 in Figure 1). This defines point E and the estimation uncertainty around it in Figure 2. Estimation uncertainty is represented in the figure as the range for E + S to E - S. Policy analyses requiring a margin of safety also then impose further risk factors such as a reduction of Mi in addition to the reduction of S in the concentration regarded as "safe". This specification for a desired (after regulation) ambient concentration determines whether the earlier decisions on how to measure the central tendency and uncertainties in health outcomes will be consequential. That is, consider the cases M1, M2 and M3 with the two different exposure profiles. My argument says that decisions about whether health outcome is serious enough to count (for regulatory purposes) and about the functions linking ambient concentrations to that outcome implicitly determine the starting point on the vertical axis -- E. Change anyone of them and we move E up or down the vertical axis. Likewise the treatment of estimation uncertainty (q) and judgments about risk factors (S) and judgments about risk factors (i.e., whether M_1, M_2 or M_3) will for a given distribution of people (AT versus AJ) yield different proportions of the population that are affected -- 0 I, 0 II, 0 III or 0 IV.

Recognizing these choices as potentially consequential implies we should be evaluating them by asking whether a count of affected people, or a cost per person, or a net aggregate benefit would be important to the regulatory decision. Each is affected in principle by decisions about what counts (and thus the E) and how uncertainty is incorporated (i.e., the S and M_i). An important element in the discussion at the workshop was that consideration should be given to the implications of making these decisions differently depending on whether the outcomes lead to dramatically different net benefits.

d) Research Complementarities Between Risk Assessment and Benefits Measurement. Table 1 summarizes a few general research tasks that were discussed at the workshop for hazardous air pollutants and the potential for complementarity or interactions that stem from how the research is designed. Three of the four tasks generally defined as involving toxicology or epidemology would benefit from complementary economic research related to specific modifications in the elements of policy analyses that could be consequential to decisions.

Table 1 Complementarities in RA[4] and BM[5] Research For Policy Associated with Hazardous Air Pollutants

TASK	RA	BM
Evaluating the nature of the physical effect on people	X	-
Evaluating the nature of the health outcome affecting people	X	Eliciting lay person's preferences for different health endpoints
Measuring distribution of ambient concentrations and number of people experiencing them	X	Evaluating the prospects for private action to mitigate or reduce exposure received
Estimating the overall consequences of Baseline (no regulation) and Regulated Alternatives	X	Defining and measuring economic choices for identified health outcomes so tradeoffs could be used to estimate benefits from policy
Evaluating importance of layperson's "worry quotient" or policy as insurance	-	Measure economic benefits as value of a regulatory program or policy

e) Parallel Research. Given the limited information, the number of hazardous air pollutants to be evaluated, and the time and resources available to develop such evaluations Lave's paper and presentation at the workshop suggested a different strategy for measuring the benefits due to HAP regulations.. He proposed that we consider measuring the economic value of the "policy" as an object of choice rather than the reductions in specified health conditions attributed to reduced ambient concentrations of individual hazardous air pollutants.[6] As noted in Table 1, this approach was discussed as a method for assessing the importance of HAP policy as providing a type of insurance for lay persons' "worries" about serious health outcomes that arise as surprises from exposures to these pollutants.

---

[4]RA refers to research in field related to risk assessment. The primary areas considered in the workshop discussion were toxicology and epidemiology. An "X" means research is clearly needed.

[5]BM refers to benefit measurement. The elements in the table illustrate research tasks that would be complementary to the task to be addressed in risk assessment

[6]This approach parallels innovations in the use of contingent valuation for the damage assessments associated with natural resource damage cases. To my knowledge it was developed for the Exxon Valdez case by Carson et al. (1992).

The workshop discussions suggested to me an opportunity for economic research to proceed along three inter-related lines that would complement each other and provide opportunities for cross validation of benefit estimates for policy. Unfortunately, there was not sufficient time to discuss this strategy at the workshop so it is not reflected in the summary of this discussion. As a result, this paper ends with a discussion of each of the three lines of activity.

The first entails a proposal implicit in Chestnut and Freeman's comments and in the summary given in Table 1. This encompasses evaluation of whether we can measure people's preferences over non-traditional health outcomes. This task involves not only rating outcomes likely to be associated with hazardous air pollutants, but also investigating the feasibility of using existing revealed preference information and stated preference surveys to recover benefit measures for these types of non-traditional choices. Agee and Crocker's (1994) study of parents' willingness to pay for reducing blood lead levels for their children is an example of the type of revealed preference analysis envisioned in this proposal. Recent applications of conjoint methods (Johnson and Desvousges [1997]) suggest it may be feasible to offer non-traditional health outcomes within this format.

The second line of research involves the focus group, survey development and pilot studies required to evaluate whether Lave's proposals to evaluate the control policy as an object of choice can actually be presented as a plausible choice alternative. It is not clear that it can, but following the protocols used in developing modern CV surveys (especially those for large scale damage assessments) it should be possible to resolve this issue without conducting a full contingent valuation study.[7]

The last line of research was presented briefly in my comments, but time did not permit it to be discussed in specific terms at the workshop. It argues for conducting the first two together because, in principle, we should be able to establish a relationship between measures of the economic value of the policy and the benefits for reducing specific health outcomes. The former is a type of ex ante option price with some private mitigation (as a type of private insurance), and the latter is a set of ex post values for the outcomes being avoided. The early logic developed by Anderson [1979] should, with modification using Graham's [1991] extension to the definition of option price, allow one to relate the two measures under specific conditions. This implies the economic value of the policy (as an object of choice) could be compared to the sum of the economic values of the avoided health outcomes. This would serve to unify the analysis, provide a check for both the CV (i.e. contingent valuation) estimate and a gauge of the potential for omission in cases where only a subset of the physical effects can be measured.

[7]A contingent valuation study rather than conjoint is proposed here because the object of choice is a policy and not a specific set of health outcomes with varying attributes. The full attributes of the results of the policy could not be described. If they could, then more conventional methods would be used. Indeed an identification of the uncertainty in the nature of the avoided health effects would likely be included as part of the description of the policy.

References

Agee, Mark D. and Thomas D. Crocker, 1994, "Parental and Social Valuations of Child Health Information", Journal of Public Economics, Vol 55 (September): 89-106.

Anderson, James E., 1979, "On the Measurement of Welfare Cost Under Uncertainty", Southern Economic Journal, Vol 45 (December): 1160-1171.

Bockstael, Nancy E. and Kenneth E. McConnell, 1999, "The Behavioral Basis of Non-Market Valuation" in Joseph A. Herriges and Catherine L. Kling, editors, Valuing Recreation and the Environment (Cheltenham, U.K.: Edward Elgar).

Carson, Richard T., Robert C. Mitchell, W. Michael Hanemann, Raymond J. Kopp, Stanley Presser and Paul A. Rudd, 1992, "A Contingent Valuation Study of Lost Passive Use Values Resulting from the Exxon Valdez Oil Spill", Report to Attorney General of State of Alaska, NRDA, Inc., November.

Graham, Daniel A., 1992, "Public Expenditures Under Uncertainty: The Net Benefits Criteria", American Economic Review, Vol 82 (December): 882-946.

Johnson, F. Reed and William H. Desvousges, 1997, "Estimating Stated Preferences with Rated-Pair Data: Environmental, Health, and Employment Effects of Energy Policies", Journal of Environmental Economics and Management, Vol 34 (September): 79-99.

Office of Air and Radiation, 1997, "The Benefits and Costs of the Clean Air Act: 1970 to 1990", U.S. Environmental Protection Agency, October.

Viscusi, W. Kip, 1997, "The Value of Risks to Life and Health", Journal of Economic Literature, Vol 31 (December): 1912-1946.